

PI Flegel WA, Wagner FF;
 XX
 DR WPI: 2001-291052/30.
 XX
 PT New nucleic acid molecular structure, useful for detection of common
 PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
 PT and RHE genes -
 XX
 PS Example 10; Fig 5; 135pp; English.
 CC The sequence represents the coding sequence of Rhesus gene locus:
 CC RHD gene deletion in Rh negative haplotypes. The Rhesus genes
 CC locus comprises the RHD, SMP1 and RHE (all undefined) genes and/or the
 CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
 CC and/or the downstream Rhesus box. The RHD and RHE genes are located at
 CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
 CC RHD deletion in the common RHD negative haplotypes. The sequence has
 CC been used to design primers which are useful for: (1) the specific
 CC detection of the common RHD positive haplotypes in D-negative
 CC individuals; (2) blood group typing; (3) determining whether a patient
 CC can be transfused with Rh negative blood and whether blood is suitable
 CC for transfusion to patients who should not be exposed to antigen C; (4)
 CC assessing the risk of a Rh negative mother of conceiving or carrying an
 CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
 CC women who are Rhesus D negative, where the foetus is not homozygous for
 CC the RHD gene to treat or prevent haemolytic disease of the newborn.
 XX
 XX Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other:
 SQ

Query Match 46.5%; Score 642.2; DB 22; Length 3093;
 Best Local Similarity 79.1%; Pred. No. 2.4e-190;
 Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

QY 352 gaataattgctcaagaagaagaatttgcctaataagaagaatttcaagcaga 411
 DB 1377 GAACCTTGAACCTTCAGAGAGATGATTACGATATCGAGAGAAATTTCAAGCAGA 1318
 QY 412 aagcaatcaagaagtgacttggtgctgttaagaagcaatcattcataagaagcaga 471
 DB 1317 AAGCATTCAGAGAGTGACTTGGTGTCTTAAAGACATTCAGTTTATTAAGGAGACAGA 1258
 QY 472 gataaagattcagaanaatttgcacctgacatgtgtataaagaanaa---ccattt 529
 DB 1257 GCATTAAGCTTGGAAAATTTGACGCTGACATGTGATGAAGAAAATTTCCATT 1198
 QY 530 tctgaagggaatcaagcttgctgcagaaatttcataatgaagaagctgaagt 589
 DB 1197 TCTGAGAGAAATTCAGCTGCTCAGAAATTTGCATGAGTAA-CAGAGCCAAATGCT 1139
 QY 590 aatcctcaagaacaatggggaataatctcctgacatgtcagaagctctcacagctcc 649
 DB 1138 AATTCCCAAGAACATGGGGAATATGCTCCAGGCATGTCAGAGGCTTATGCGCAACC 1079
 QY 650 atcaataactgctgtagagcct-aggagaataatgttttctgtagacagccaggttc 708
 DB 1078 CTCCTCATCAGAGTCCAGAGGATATCAGGAAAAAATGTTTGTGGCCAGCGCGGTC 1019
 QY 709 cctgtgctgtgtagcagcctagagacttggtgctgtgtccagtgtccagta----- 755
 DB 1018 CTCATGCTGTGTGACGCTCAGGACTGTGCTGCTGATCCACGACATCCCAACCATTA 959
 QY 756 -----attcagctgtgtctcagaaggtgcagaagccccaagcctt 794
 DB 958 CTGACGGAGGAGCAAGTATGAGCTTGGCTGTAGCTTCCGGGAGTGCAGGCCCAAGCCTT 899
 QY 795 ggcacactcccaagtgtgttgagcctgtggtgcagaagaatgcagaatgtgagttgg 854
 DB 898 GACACCTTCCAGTGTGTGTGAGACTGCGACAGCAAGCAAGTAATGGGGTTTGA 839
 QY 855 aaccccaatcagattcagaagataatagaaaccccgagtcagccaggaagtgtg 914
 DB 838 AACCTTCCGCTTGAATTAAGAGAGATGTGCGGAAATGCTTGATGCCCAAGAGTTTG 779

QY 915 ctgtaagggtggtgtcctcatatggagaacctctgcagaaggtagtagaanaagggaatgttg 974
 DB 778 CTGACAGAGACAGAGGCCCTCATGAGATCCTGTGCGAGGCGAGTGCAGAAAGGAATGTGG 719
 QY 975 ggttgagagcccccacacagatgccagtggtggtcctcatctagtagactgtgagaaga 1034
 DB 718 GGTCAAGACACCCACACAGTCCCTACTGCGGACACACCTAGTGTGAGCAAGAG 659
 QY 1035 gtccacatccctccagaagcctcagaaggttagatccactacagctgtgcagctgtgctg 1094
 DB 658 G-----TCTTCAGACCCCAAGATGTAGATCCACCAAGACTTGACACCTGTACTG 606
 QY 1095 aaaaatccacagacactcagtgccagctgtgaaagcagcaggaatgtagtctgacct 1154
 DB 605 GAAAACTCAGACACTCATATGCGACACCATGAAACAGACTGAGAGAGGCGTACCT 546
 QY 1155 aaaaaacgtatgtagagagctgcagcaaacgctgtggaatctacacctgtgactgtgcatg 1214
 DB 545 GCAAAAGTACAGGGGAGAGCTGCCCAAGACCATGGGAACCCCTTCATCTGCGTG 486
 QY 1215 acctgagctgtagacatgtagatcaaaagagatcatttggagcttaagattgactgccc 1274
 DB 485 ACTGATGTGAGATGTGAGTCAAGAGATCATTTTGGAGCTTTAAAGATTGACTGCC 426
 QY 1275 caactggaattcggactatattggtggtccgtacacctgttggccaattttccat 1334
 DB 425 CCAGTGAATTTGAGACTCATGAGGCGCTGTAGCGCTCTTGTGTTGACCAATTAATCCAT 366
 QY 1335 ttggaactgcgtattaccacatgcctgtacctcatatgtatgtag 1381
 DB 365 TTGGAATGTGCTGATTAATTAACCAATGCTGTACCCCATTTGATTAATCTAG 319

RESULT 3
 AAS03689/C
 ID AAS03689 standard; DNA; 9236 BP.
 XX
 XX AAS03689;
 AC
 XX
 XX
 DT 29-AUG-2001 (first entry)
 DE Rhesus gene locus: upstream Rhesus box of D-positives.
 KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
 KW RHE; Rh negative; blood group typing; blood transfusion; antigen C;
 KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
 OS Homo sapiens.
 XX
 XX
 PN WO200132702-A2.
 XX
 PD 10-MAY-2001.
 XX
 PF 31-OCT-2000; 2000WO-EP10745.
 XX
 PR 02-NOV-1999; 99EP-0121686.
 PR 31-MAY-2000; 2000EP-0111696.
 XX
 PA (DRB-) DRK BLUTSPENDEDIENST BADEN WÜRTTEMBERG.
 XX
 PI Flegel WA, Wagner FF;
 DR WPI: 2001-291052/30.
 XX
 PT New nucleic acid molecular structure, useful for detection of common
 PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
 PT and RHE genes -
 XX
 PS Disclosure; Fig 9; 135pp; English.
 CC The sequence represents the coding sequence of Rhesus gene locus:
 CC upstream Rhesus box of D positives. The Rhesus genes locus

CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with Rhd negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a Rhd negative mother of conceiving or carrying an
CC RHD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.

XX Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match 46.5%; Score 642.2; DR 22; Length 9236;
Best Local Similarity 79.1%; Pred. No. 4.5e-190;
Matches 844; Conservative 0; Mismatches 178; Indels 45; Gaps 5;

QY 352 gaatttgcctaagagaagaacaatttggcctaagaagaatttctaagacga 411
DB 6053 GAACCTTGAACCTTCAGAGAGATGATTTAGACTATCTGCGAGAGAAATTTCTAAGACGA 5994
QY 412 aagcattcaagaagtgactgggtgtgttaagaagcattcaattcataaggaagcaga 471
DB 5993 AACCATTTCAAGAGGTGACTTGGTGCTGTAAAGACATTCAGTTTATTAAGGAGACAGA 5934
QY 472 gcatatgattcagaatttgcacccctgacatgtatataaaaaaa---ccattt 529
DB 5933 GCATTAAGATTCGAAAATTTGCAACCTGACATGTATGAAAAGAAAATTTCCATT 5874
QY 530 tctgaagggaattcaagctgctgcagaaattgcataatgaatgaagctgaatt 589
DB 5873 TCTGAGAGAAATTCAGCTGGCTGCAGAAATTTGCATGATGA---CAGAGACCAATGCT 5815
QY 590 aatcctcaagaacaatgagggaataatctctgcacatgtcagaagctctcaagcaatgc 649
DB 5814 AATTTCCAGACAAATGGGGAAAATGCTCTCAGGGCATGTAGAGAGCTTTATTCAGCAACC 5755
QY 650 ataaatcactggcctgtagagcct-aggagaagaatgtttgaggagagcccaaggttc 708
DB 5754 CTCCCATCAACAGAGTCAGAGATAGGAAAAAATGTTTGTTCGCGAGCCCGGGGCTC 5695
QY 709 cctgtgtgtgtgcagccttagaagactgtgtccctgtgtccagta----- 755
DB 5694 CTCATGCTGTGTGACAGCGCTAGGAGACTTGCTCCCTGCATCCAGCCACATCCCAACCATGA 5635
QY 756 -----atcagctgtgcttcagaaggtgtagaagcccaagcctt 794
DB 5634 CTGACGGGAGGAGATGAGACTTGGGCTGTAGCTTGGGGAGGTGCAACCCCAAGCCCTT 5575
QY 795 ggcagcttcaagatgtgttagcctgtggtgtgcaagaagtaagaatgtgagtttgg 854
DB 5574 GACAGCTTCCATGTGTGTGTAGACTGTGAGAGTGCACAGAAAGTAAATTTGGGGTTTGA 5515
QY 855 aacctcaatcagattcagaagatataatggaacccctgtgattcccaagcagcagtttg 914
DB 5514 AACCTTCGCGCTAGATTAAGAGAGATGCGGAATTCCTGGATGCCAGTCCAGAGATTGG 5455
QY 915 ctgtgaagggtgtgtcctcaatgagaacctctgcagaaggtagtacaaaggggaatgttg 974
DB 5454 CTTCAGAGAGACAGGCGCTCATATGAGATCTTCCAGGCGCAGAGAGGGAATTTGTGG 5395
QY 975 ggtgaggagccccaacacagagtcccccagtggtgtcatttaataagctgtgagaaga 1034
DB 5394 GGTCAAGAGACCCACACACAGATCCCTACTGTGGGACACCACTTAATGAGAGAGAG 5335
QY 1035 gtccacacatccccaagcctcagaaggttaatcactgacactgtgcagcagctgtgcttg 1094
DB 5334 G-----TCTTCAGAGACCCACAGATGTGTATCTCACCGACAGCTTGACCCGTTACTG 5282

QY 1095 aaaaaccacagacactcaatgctccagcctgtgaaagcaagaagga tggagctgttacct 1154
DB 5281 GAAAGCTGACAGACCTCAATGCGAGCCCATGAAGCAGCTGAGAGGGAGGCTGACCT 5222
QY 1155 acaaacctgtgtgcagagctgcagcaagacgctgtggaattcaactcttcagctgtcatg 1214
DB 5221 GCAAAGGTACAGGGGACAGCTGCCAAGACATGGGAACCCACCCCTTCATCGCTG 5162
QY 1215 acctgagcttagagacatgagatgaaagagatcattttagagcttaagattgactgac 1274
DB 5161 ACCTGAGATGTAGATGTGAGACTCAAGAGAGATCATTTTGGAGCTTAAAGATTGACCTCC 5102
QY 1275 ccactgattcgcagactatattggggccgtaacccttgtttggccaatttttccat 1334
DB 5101 CCAGTGGATTTTGGACCTCATGAGGCGCTGTAGCCTCTTGTGTTGACCAATTTATCCAT 5042
QY 1335 ttggaagctccgattatcccaatgcctgtacttcattgtatgag 1381
DB 5041 TTGGAATGGCTGTATTATCCCAATGCTGTACCCCATTTGTATCTAG 4995

RESULT 4
AAS03690/C
ID AAS03690 standard; DNA: 9236 BP.
XX
AC AAS03690;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: downstream Rhesus box of D-positives.
XX
XX Rhesus box; RHD positive; sequence length polymorphism; SSP: RHD; SMP1;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS Homo sapiens.
XX
PN WO200132702-A2.
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000WO-EP10745.
XX
PR 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTEMBERG.
XX
PI Flegel WA, Wagner FF;
XX
XX WPI; 2001-291052/30.
DR
XX
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PS Disclosure; Fig 10; 135bp; English.
XX
XX The sequence represents the coding sequence of Rhesus gene locus:
CC downstream Rhesus box of D positives. The Rhesus genes locus
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with Rhd negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a Rhd negative mother of conceiving or carrying an
CC RHD positive foetus. Anti-D antibodies are useful for treating pregnant

Db 6053 GAACCTTGAACTTCAGAGAGATGATTTAGAGTACTGGCAGAGAGAAATTTCTAACGACGA 5994
Qy 412 aagcatcaagaagtgacttggtgctgtataaaggcatcactcaataaaggagcaga 471
Db 5993 AAGCAATTCAGAGAGTGGATTTGGTGTAAAGACTTACAGTTATATAGGAGGACGA 5934
Qy 472 gcaataagatccaagaatttcacacctgacaatgtagataaanaaanaa--cccatc 529
Db 5933 GCATTAAGATTCGAAATTTGGACCTCGACAAATGATAGAAAATAATTTCCCATTT 5874
Qy 530 tctgaggggaaattcaagctgctgcagaatttgcataatgtaagagagcagtaatt 589
Db 5873 TCTGAGGAGAAATTTACACTGGCTGCAGAAATTTGCATGAGTAA--CAGAGCCAAATGCT 5815
Qy 590 aatcctcaagaacaatgaggaaataatcctcggacatgltcaagaggtctcacagagtc 649
Db 5814 AATTCGCAAGCAATGGGGAATGTCTCCAGGGCATGTCCAGAGGTCTTATGGCAACCC 5755
Qy 650 atcaaatcaagctgagagcct--aagagaaaaatggttttgtggagacagccaggtc 708
Db 5754 CTCCTTCACAGAGTCCAGAGGTATCAGGAAAAATGATTTGTTGGCCAGGCCGGGGTTC 5695
Qy 709 cctgtgtgtgtgacagcttagagactgtgtccctgtgtccagta----- 755
Db 5694 CTCATGCTGTGTGACGCTTAGGACCTGTGCTCCCTGCATCCAGCCATCCCAACCATGA 5635
Qy 756 -----atcaagctgtggtctcaagaggtgcaagcccaagcct 794
Db 5634 CTCACGGAGGCAAGGTAGAGCTTGGGCTGTAGTCTCGGGAGTGCACGCCCAAGCCCTT 5575
Qy 795 ggaagcttccaagtggtgttagcctgtgggtgcaagaagtcaaaatgaggtttgg 854
Db 5574 GAAAGCTTCAGATGTGTGTAGACTGAGTGCAGCAAGAAATGAGTGGGTTTGA 5515
Qy 855 aacctccaatcagatcattcaagatatatagaaacccctgatatcccaagcagaattgg 914
Db 5514 AACCTTCGGCTAGATTAAAGAGATGTGGCGAAATGCTGTGATGCCCATCGAAGTTTG 5455
Qy 915 ctgtagagggtggtgtcctcatgtagaactctgcgaaggtagtagaanaaaggaatgtg 974
Db 5454 CTCGACGAGACAGGCCCTCATAGGAGATCTCTGCGCAGGCGAGTGCAGAGGGAATGTGG 5395
Qy 975 ggtgaggagccccaacagatcccccgtgggtgcattagtagagctgtagaagaa 1034
Db 5394 GGTCAAGAGACCCACACACAGTCCCTACTGGGGCACCACTAGTGAGAGAAAG 5335
Qy 1035 gtccacacatccctcagactccaagaaggtagatccacttgcagactgtgacctg 1094
Db 5334 G-----TCTCCAGAGACCCCAAGATGTGTAGATCCACCGACACTTGTAACCTG 5282
Qy 1095 aaaaatccacagacacatcaagtgcagcctgtgaaagcagcagagatgagctgacct 1154
Db 5281 GAAAAAGCTGCAGACTCAATCAATGCCAGCCCATGAAGAGAGCTGAGAGGCGTAACTT 5222
Qy 1155 acaaaacgttagtgcaagctgtagcacaagacgttgggaatcctcctcgtgcatgt 1214
Db 5221 GCAAAAGTACAGGGGCAAGGCTGCCCAAGACCAAGGGAACCCCTTCATCTGCGGTG 5162
Qy 1215 acctgaagctgagacatgtagtcaaaagagatacttttggagctttaaagattgactcc 1274
Db 5161 ACCTGGATGTGAGATGTGGAGTCAAAAGAGATCATTTTGGAGCTTTTAAATTTGACTGCC 5102
Qy 1275 ccactgatttcgactatatatggggcctgtaccacctgttttggccaattttlcat 1334
Db 5101 CCACCTGGATTTTGGACTCTCATGAGGCGCTGTAAGCTCTTTTGTGGCAATTTATCCCAT 5042
Qy 1335 ttggaactgcgttatttaccacatgctgtactcactgtatgtag 1381
Db 5041 TTGGAATGCTGTATTTACCAATGCGTAAACCCCATTTGTATCTAG 4995

RESULT 6
AAU03635

ID AAU03635 standard; DNA: 6063 BP.
XX
AC AAU03635;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6123.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01339.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232387.
PR 14-SEP-2000; 2000US-0232398.

[illegible]

QY 957 tacaaagggaatgttgggtggagccccacacagatccccagttgggtccatcta 1016
D 4079 t-gtgaagggaatgttgggtggagccccacacagatccccagttgggtccatcta 4137
QY 1017 gttagactgttgaaagagtcaccatctccagactccagaggaagtaatactactgaca 1076
D 4138 gttagactgttgaaagagtcaccatctccagactccagaggaagtaatactactgaca 4197
QY 1077 gcttcagagatgtgcctgaataatccacagacactcagctgcagctgtgaagacagac 1136
D 4198 gcttcagagatgtgcctgaataatccacagacactcagctgcagctgtgaagacagac 4257
QY 1137 gga-tggagtcctgtaccctacaaacgtagtggcagagctgcacagacgctgggaac 1195
D 4258 gga-tggagtcctgtaccctacaaacgtagtggcagagctgcacagacgctgggaac 4317
QY 1196 tactcttcgtcatgtgacgtgagcgtgagacatgagtaacaaagatatttttga 1255
D 4318 tactcttcgtcatgtgacgtgagcgtgagacatgagtaacaaagatatttttga 4377
QY 1256 gcttgaagtgtgacgtccacactggatctgcagctataatggggccgt-acccttg 1314
D 4378 agcttaataattgactgctgcgtgagatttggacttgacatggggtttgaacgcttg 4437
QY 1315 ttgtgccaatttttccattggaactgcgcgtatttaccatgaccttaccatg 1374
D 4438 ttgtgccaatttttccattggaactgcgcgtatttaccatgaccttaccatg 4497
QY 1375 tatgtag 1381
D 4498 tatgtag 4504

RESULT 7

AAD05134
ID AAD05134 standard: cDNA: 3203 BP.

AC AAD05134;

DT 17-JUL-2001 (first entry)

XX Human secreted protein-encoding gene 14 cDNA clone H15BF60, SEQ ID NO:24.

XX Human: secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiotensin disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnary;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Key location/Qualifiers

XX CDS 153..356

XX sig_peptide /tag= a

XX mat_peptide /product= "Human secreted protein"

XX MO200134769-AZ.

XX 17-MAY-2001.

XX 01-NOV-2000; 2000MO-US30040.

XX 05-NOV-1999; 9905-0163580.
PR 30-JUN-2000; 2000US-0215130.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
PI WPI; 2001-308781/32.
DR P-PSDB; AAE01245.
XX New isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PS Claim 1; Page 408-409; 519pp; English.

AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the presence of
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 24 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.

XX Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;

Query Match 44.9%; Score 620.4; DB 22; Length 3203;
Best Local Similarity 78.3%; Pred. No. 1.7e-183;
Matches 835; Conservative 0; Mismatches 192; Indels 40; Gaps 6;

QY 352 gaatttgggtgaatggaggaacaaatttgggtatataaggaagaattttaaagcaga 411
D 1599 gaatttgggtgaatggaggaatgatttgaaggtatctggtcgaagaattttaaagcaga 1658
QY 412 aagcatcaagagtgacttgggtgctgtaaaagcattccatcaaaaggagcaga 471
D 1659 aagcatcaagagtgacttgggtgctgtaaaagcattccatcaaaaggagcaga--aaca 1716
QY 472 gataaggttcaagaatttgcacccgtgacatgtgataaaagaagaacccatttc 531
D 1717 gataaagaatttggaaaatttgcacccagtgctgacgaagaagagaaacccgttttt 1776
QY 532 tgaagggaattcaagctggtcgcagaatttgcataatgtaaaaggagcgtgaattaa 591
D 1777 tgaagggaattcaagctggtcgcagaatttgcataatgtaaaaggagcgtgaattaa 1836
QY 592 tctcaagaacaatggggaataatctccctgagacatgtcagaagcttcaagcgctcat 651
D 1837 tctcaagaacaatggggaataatgtctccagagcatgtctacaggtcttcaagcgccct 1896

OY	652	caataactgcgcctggagggccttagag---aaatggttttctgtggacagcccaagggtc	708
Db	1897	cccatcacagacccttgaaagactagagaaataatggtttctgtggccaggcccaagggtc	1956
OY	709	ccctgtgcctgtgtgagagcttagagacttgtagccctgtgtcccaagttaatctag-----	761
Db	1957	cccatgtcctgtgtgagccttgtagaacttgtagccctgtgtcccaagctgtctccagcaatgyc	2016
OY	762	-----ctgtgacttcagaggtgtgcaagccccaagccttgg	796
Db	2017	taaaaggccgaggtatagactctgtgcccatgtttcttagaggtgtgcaagccccaacttgg	2076
OY	797	cagcttccaaagtgtgtcttgagcctgtgtgtgtgcaagaagtcaagaattgagtttggaa	856
Db	2077	cagcttccaaagtgtgtcttgagcctgtgtgtgtgcaagaattcaagaattgagtttggaa	2136
OY	857	ccctcaatcagaattctcagaagatatagaaaccccttgatgcccagagaaagtttct	916
Db	2137	ccctcaatcagaattctcagaagatatagaaaccccttgatgcccagagaaagtttct	2196
OY	917	gtagagctgtggtgtctcatgtgaaagcctctgtgcaaggtgtgtacaaaagggaaatgttgg	976
Db	2197	cgaggggcagagcctctcatgtgaaagcctctgttagtgaggt-gtgaagggaaatgttggg	2255
OY	977	tggagagcccccacacaaagatcccccagtgtggggccctcatctagtagagctgttggaaagagt	1036
Db	2256	tggagagcccccacacaaagatcccccagtgtggggccctcatctagtagagctgttggaaagag	2315
OY	1037	ccacatctctcacaagctccagaagggtatgattcaatgcagagccttgacagatgtgcctgaa	1096
Db	2316	ccacagctctctcagaacccccaagatgtgtatgattccacgcagagccttgtagccgtgtcaccttga	2375
OY	1097	aaatccacaagacacttcagctgtgccagcctgttgaagacag-cagggaatgtagtctgttacccta	1155
Db	2376	aaagccaacaagccctaaacgcagcccggtgaagacagcaggaggtgggagcctataaccctg	2435
OY	1156	caaaaacgttagtgtgcagaagctgtgacaaagacccgtgtggaaatctactcctctgtcatgtcaga	1215
Db	2436	tgaagcccaaggggcagaagctgtgcccagaagactaaggagaaacctactcctctgtcatcatgtga	2495
OY	1216	ccctgcagctgtgagacatgtagtcaaaagagatacttcttgagccttaagatttgtagtgcct	1275
Db	2496	ccctgcagctgtgagacatgtagtcaaaagagatacttcttgagccttaagatttgtagtgcct	2555
OY	1276	cacttgatcttcggaactatatactggggcccgf-accctctgttcttgagcccaatttttccat	1334
Db	2556	cgcctgatttttgagctctgcaatggggcctgtgaaacgccttctgttggcaggtttccccat	2615
OY	1335	ttggaactgcgcgtatttaccacaatgcctgtgactccatctgatatgtag	1381
Db	2616	ttggaactgcgtatttaccacaatgcctgtgactccatctgatatgtag	2662
RESULT 8			
ID AAL03634 standard; DNA: 6063 BP.			
XX	AAL03634:		
XX	21-NOV-2001 (first entry)		
XX	Human reproductive system related antigen DNA SEQ ID NO: 6322.		
XX	Human: reproductive system related antigen; reproductive system disorder		
XX	cancer; gene therapy; ds.		
XX	Homo sapiens.		
XX	WO200155320-A2.		
XX	02-AUG-2001.		
XX	17-JAN-2001; 2001MO-US01339.		
XX	PF		

XX	PR	31-JAN-2000	2000US-0179065
PR	04-FEB-2000	2000US-0180628	
PR	24-FEB-2000	2000US-0184664	
PR	02-MAR-2000	2000US-0186550	
PR	16-MAR-2000	2000US-0189874	
PR	17-MAR-2000	2000US-0190076	
PR	18-APR-2000	2000US-0198123	
PR	19-MAY-2000	2000US-0205515	
PR	07-JUN-2000	2000US-0209647	
PR	28-JUN-2000	2000US-0214886	
PR	30-JUN-2000	2000US-0215135	
PR	07-JUL-2000	2000US-0216647	
PR	07-JUL-2000	2000US-0216880	
PR	11-JUL-2000	2000US-0217487	
PR	11-JUL-2000	2000US-0218296	
PR	14-JUL-2000	2000US-0218790	
PR	26-JUL-2000	2000US-0220964	
PR	26-JUL-2000	2000US-0220964	
PR	14-AUG-2000	2000US-0224518	
PR	14-AUG-2000	2000US-0224519	
PR	14-AUG-2000	2000US-0225213	
PR	14-AUG-2000	2000US-0225214	
PR	14-AUG-2000	2000US-0225266	
PR	14-AUG-2000	2000US-0225266	
PR	14-AUG-2000	2000US-0225267	
PR	14-AUG-2000	2000US-0225268	
PR	14-AUG-2000	2000US-0225270	
PR	14-AUG-2000	2000US-0225447	
PR	14-AUG-2000	2000US-0225757	
PR	14-AUG-2000	2000US-0225758	
PR	14-AUG-2000	2000US-0225759	
PR	18-AUG-2000	2000US-0226279	
PR	22-AUG-2000	2000US-0226681	
PR	22-AUG-2000	2000US-0226688	
PR	22-AUG-2000	2000US-0227182	
PR	23-AUG-2000	2000US-0227189	
PR	30-AUG-2000	2000US-0228924	
PR	01-SEP-2000	2000US-0229287	
PR	01-SEP-2000	2000US-0229343	
PR	01-SEP-2000	2000US-0229344	
PR	01-SEP-2000	2000US-0229345	
PR	05-SEP-2000	2000US-0229509	
PR	05-SEP-2000	2000US-0229513	
PR	06-SEP-2000	2000US-0230437	
PR	06-SEP-2000	2000US-0230438	
PR	08-SEP-2000	2000US-0231242	
PR	08-SEP-2000	2000US-0231244	
PR	08-SEP-2000	2000US-0231244	
PR	08-SEP-2000	2000US-0231413	
PR	08-SEP-2000	2000US-0231413	
PR	08-SEP-2000	2000US-0231414	
PR	12-SEP-2000	2000US-0231968	
PR	14-SEP-2000	2000US-0232397	
PR	14-SEP-2000	2000US-0232398	
PR	14-SEP-2000	2000US-0232399	
PR	14-SEP-2000	2000US-0232400	
PR	14-SEP-2000	2000US-0232401	
PR	14-SEP-2000	2000US-0232401	
PR	14-SEP-2000	2000US-0233063	
PR	14-SEP-2000	2000US-0233064	
PR	14-SEP-2000	2000US-0233065	
PR	21-SEP-2000	2000US-0234223	
PR	21-SEP-2000	2000US-0234224	
PR	25-SEP-2000	2000US-0234997	
PR	25-SEP-2000	2000US-0234998	
PR	26-SEP-2000	2000US-0235484	
PR	27-SEP-2000	2000US-0235834	
PR	27-SEP-2000	2000US-0235834	
PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	

PR	02-OCT-2000	2000US-0236802	PR
PR	02-OCT-2000	2000US-0237037	PR
PR	02-OCT-2000	2000US-0237038	PR
PR	02-OCT-2000	2000US-0237039	PR
PR	02-OCT-2000	2000US-0237040	PR
PR	13-OCT-2000	2000US-0239935	PR
PR	13-OCT-2000	2000US-0239937	PR
PR	20-OCT-2000	2000US-0240960	PR
PR	20-OCT-2000	2000US-0241221	PR
PR	20-OCT-2000	2000US-0241785	PR
PR	20-OCT-2000	2000US-0241786	PR
PR	20-OCT-2000	2000US-0241808	PR
PR	20-OCT-2000	2000US-0241809	PR
PR	20-OCT-2000	2000US-0241826	PR
PR	01-NOV-2000	2000US-0244617	PR
PR	08-NOV-2000	2000US-0246474	PR
PR	08-NOV-2000	2000US-0246476	PR
PR	08-NOV-2000	2000US-0246477	PR
PR	08-NOV-2000	2000US-0246478	PR
PR	08-NOV-2000	2000US-0246523	PR
PR	08-NOV-2000	2000US-0246524	PR
PR	08-NOV-2000	2000US-0246525	PR
PR	08-NOV-2000	2000US-0246526	PR
PR	08-NOV-2000	2000US-0246527	PR
PR	08-NOV-2000	2000US-0246528	PR
PR	08-NOV-2000	2000US-0246532	PR
PR	08-NOV-2000	2000US-0246609	PR
PR	08-NOV-2000	2000US-0246610	PR
PR	08-NOV-2000	2000US-0246611	PR
PR	17-NOV-2000	2000US-0249207	PR
PR	17-NOV-2000	2000US-0249208	PR
PR	17-NOV-2000	2000US-0249209	PR
PR	17-NOV-2000	2000US-0249210	PR
PR	17-NOV-2000	2000US-0249211	PR
PR	17-NOV-2000	2000US-0249212	PR
PR	17-NOV-2000	2000US-0249213	PR
PR	17-NOV-2000	2000US-0249214	PR
PR	17-NOV-2000	2000US-0249215	PR
PR	17-NOV-2000	2000US-0249216	PR
PR	17-NOV-2000	2000US-0249217	PR
PR	17-NOV-2000	2000US-0249218	PR
PR	17-NOV-2000	2000US-0249244	PR
PR	17-NOV-2000	2000US-0249245	PR
PR	17-NOV-2000	2000US-0249246	PR
PR	17-NOV-2000	2000US-0249265	PR
PR	17-NOV-2000	2000US-0249267	PR
PR	17-NOV-2000	2000US-0249268	PR
PR	17-NOV-2000	2000US-0249269	PR
PR	17-NOV-2000	2000US-0249270	PR
PR	17-NOV-2000	2000US-0249271	PR
PR	17-NOV-2000	2000US-0249272	PR
PR	17-NOV-2000	2000US-0249273	PR
PR	17-NOV-2000	2000US-0249300	PR
PR	01-DEC-2000	2000US-0250160	PR
PR	01-DEC-2000	2000US-0250391	PR
PR	05-DEC-2000	2000US-0251030	PR
PR	05-DEC-2000	2000US-0251988	PR
PR	05-DEC-2000	2000US-0256719	PR
PR	06-DEC-2000	2000US-0251479	PR
PR	08-DEC-2000	2000US-0251856	PR
PR	08-DEC-2000	2000US-0251868	PR
PR	08-DEC-2000	2000US-0251869	PR
PR	08-DEC-2000	2000US-0251989	PR
PR	08-DEC-2000	2000US-0251990	PR
PR	11-DEC-2000	2000US-0254097	PR
PR	05-JAN-2001	2001US-0259678	PR
PA	(HUMA-) HUMAN GENOME SCI INC.		PA
XX			XX
PI	Rosen Ca, Barash SC, Ruben SM;		PI
XX			XX
DR	WPI; 2001-465570/50.		DR
XX			XX
PT	Isolated nucleic acid molecule encoding a reproductive system antigen		PT
is	used in preventing, treating or ameliorating a medical condition -		

XX	Disclosure; SEQ ID NO 6322; 1297pp + Sequence Listing; English.
PS	
CC	The present invention provides the protein and coding sequences of a
CC	number of human reproductive system related antigens. These can be used
CC	in the prevention and treatment of reproductive system disorders,
CC	including cancer. The present sequence is a genomic sequence encoding a
CC	protein of the invention.
XX	
XX	Sequence 6063 BP; 1656 A; 1163 C; 1262 G; 1982 T; 0 other;
XX	
Query Match	44.9%; Score 619.8; DB 22; Length 6063;
Best Local Similarity	78.3%; Pred.No. 3.7e-183;
Matches 835; Conservative	0; Mismatches 192; Indels 40; Gaps 6
QY	352 gaatttgcgtaaatgagagaagcaaatattggcgtataaaggaagaatctctaagcaga 411
DB	3441 gaacttgcattcttgagagagatgatttaagtgatctgcgtcgaagaatctctaagcaga 3500
QY	412 aagcattcaagagtgagcttggtgtcgtttaaaggcattcagttcataaaggagagcaga 471
DB	3501 aagcattcaagagtgagcttggtgtcgtttaaaggcattcattcataaaggaga--aac 3558
QY	472 gcaataaggttcaaaaattgcaccctgcacatctgtgataaagaagaaccattttc 531
DB	3559 gataaaaaatttgaaaaatttgaccctgtatgacgcagaagaagaagaaccctgtttt 3618
QY	532 tgaaggggaattcaagcttgctgcagaaaaattgtcatatgtaatgagagagctgaa 591
DB	3619 tgaaggagaattcaagcttgctgcagaaaaattgtcataagttacaagggagctgaa 3678
QY	592 tctccaagaacatgagggaataatctcccgagcagttaaggggtcttcacagcgctcac 651
DB	3679 tctccaagaacatgagggaaaaatgctcccgagcagtcacacagctctcaacagcagccct 3738
QY	652 caaatcaactggtccctggaggtcctagag--aaaatggtttgttgagacagggcccaagggtc 708
DB	3739 cccatcacagaccctgagagactagagagaaaaatggtttgttggtccagggcccaagggtc 3798
QY	709 cctgtgctgtgttgtagagcttagaacttggctgcctgtgccagtttaatcag----- 761
DB	3799 cccatgtgtgtgtgtagagcttagaacttggctgcctgtgccacccagctgtccacgattgc 3858
QY	762 -----ctgtggtcttaagagggtgtgcaagccccaagcccttg 796
DB	3859 taaaagggcagagtgtaagctctgcgcgcggtgttcaagggggtgcaagccccaaccttg 3918
QY	797 cagcttccaagtggtgttgagcctgtgggtgtgcaagaagtcgaagatgagtttggaa 856
DB	3919 cagcttccaatgtgtgttgagcctgtgggtgtgcaatgaaatgagatgagtttggaa 3978
QY	857 cctccaacacagatctcagaagatatgtgaaaaacccttgatgcccacagagaattgtgc 916
DB	3979 cctccatctgattctcagaagatgatatgttgaataccacttgatgcccacagcaaaagtgtgc 4038
QY	917 gtaggggtgtggtgtcctcatgagagaacctctgtcaagggtagtatacaagaagggaatg 976
DB	4039 ccaagggtgagagccttcatgtgagaacctctgtctlagggcagt-gtgaagggaagaatgt 4097
QY	977 tgggaagccccaacagaaggtcccccagtggtgttcatactaatgaaagttgtgagaagaagt 1036
DB	4098 ttggagccccaacagaataccttacttgggtggaaccaactaagtggagctgtgtgtgcaagaagg 4157
QY	1037 ccacatctctcagaactccagaaggggtatgatacacttgcaactgtgcagcattgtgcgaa 1096
DB	4158 ccacgttctctcagaaccccaagatgtgtlagaatccacgcagcagctgttgcgtgcacgtgga 4217
QY	1097 aaatccacagacactcaggtgcagcctgtgtgaaagag--cagagatgagatctgtaaccta 1155
DB	4218 aaaggtcacaagcccttaaacgcagcccggtgaagaagcagaagatgttgggtactaacctgt 4277
QY	1156 caaaacgtataatgcaaaagcttgaacgaagccgttggaatcttcaactctgtcaattctaa 1215

Db 4278 tgaagcccaaggagagcctgcccagaactaaggaaacctcctcttgcattcgtga 4337
Oy 1216 cctggagctggaacatgagatcaaaagatcatttggagcttgaagctgagctcc 1275
Db 4338 cctggaatgtagactatgagatcagagagatcatttggaaagtataatgagctgct 4397
Oy 1276 cactgagcttgcagctatatagggcccgt-acccctgttggccaattttccat 1334
Db 4398 cgtgagcttgcagctatgagctgagccttggcttggcgacatttcccat 4457
Oy 1335 ttggagcgcgctatttcccaatgctgctgactcattgtagag 1381
Db 4458 ttggaatgctgagcttacccttaccctgcttcccatgtatgag 4504

RESULT 9
ABA90193
ID ABA90193 standard; DNA: 465237 BP.
XX ABA90193;
AC ABA90193;
DT 11-FEB-2002 (first entry)
XX
DE Human oestrogen receptor alpha gene.
XX
KW Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syne-2;
KW synaptic nuclei expressed gene 2; haplotype; cytosolic; osteopathic;
KW cardiac; vasotrophic; gene therapy; vaccine; cancer; osteoporosis;
KW cardiovascular disease; oestrogen receptor; ds.
XX
OS Homo sapiens.
XX
PN WO200162969-A2.
XX
PD 30-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05358.
XX
PR 22-FEB-2000; 2000US-183756P.
PR 20-OCT-2000; 2000US-0692414.
PR 24-JAN-2001; 2001US-0768184.
XX
PA (PEKE) PE CORP NY.
XX
PI Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR WPI: 2002-041152/05.
DR P-PSDB: AAG68251.
XX
PT Novel variant of estrogen receptor alpha polypeptide useful for
PT determining the biological activity of a protein for high throughput
PT screening and for raising antibodies that elicit an immune response in
PT host -
XX
PS Example; Fig 1 page 1-93; 333pp; English.
XX
XX The present invention describes an isolated peptide (I) consisting of an
XX amino acid sequence selected from: (a) the amino acid sequence of a
XX variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
XX or (b) a fragment comprising at least 10 contiguous amino acids of the
XX protein in AAG68251. (I) has cytostatic, osteopathic, cardiac and
XX vasotrophic activities, and can be used in gene therapy and vaccine
XX production. (I) is useful for identifying an agent that binds to (I), by
XX contacting (I) with an agent and assaying the contacted mixture to
XX determine whether a complex is formed with the agent bound to the
XX peptide. A polynucleotide (II), encoding (I), is useful in the
XX development of diagnostics and therapies for diseases and disorders
XX mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
XX gene therapy for treating cancer, osteoporosis and cardiovascular
XX diseases. The human ESR-alpha gene is located on chromosome 6. The
XX present sequence represents the human ESR-alpha gene, which is given in
XX the exemplification of the present invention.

XX
SQ Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 4 other;
Query Match 44.9%; Score 619.6; DB 24; Length 465237;
Best Local Similarity 78.2%; Pred.No.5.6e-182;
Matches 849; Conservative 0; Mismatches 194; Indels 43; Gaps 7;
Oy 335 ctgcattatctgcaagctgataatttggcctaagagaaagcaatttgcataaagaa 394
Db 438730 cctgctgtagagatttggtaatttgaacttggagagaatattgggtgactcgttaaa 438789
Oy 395 gaaattcttaagcagcaaaagcatcgaagagtgacttgggtgtctttaaagcatcagt 454
Db 438790 gaaattcttaagcagcaaaagcatcgaagagtgacttgggtgtctttaaagcatcagt 438849
Oy 455 ttcatagaaggagcagagatagaagttcagaataattgcaccctgagacatgtgataaa 514
Db 438850 tt--tataagagaacaagataaagttcagaataattgcaccctgagatgagcagtagaa 438907
Oy 515 aagaaaaaaccca-tttctcgaaggagaaatcagaactgctgcagaatctgcatatgtaa 573
Db 438908 aagaaaaaacccatttttggagagagaatcagaactgctgcagaatctgcatatgtaa 438967
Oy 574 tgaagagctgaattgttaattcctcaagacaaatgggaaatactcctcagacatgcagag 633
Db 438968 caagagccaataatgttaattcctcaagacaaatgggaaataatgtctcagaagcatgtag 439027
Oy 634 gctctcaagcagctcattcaaatcactgctcggagggccttaagag-aaaatgtttt 689
Db 439028 gctctcaagcagcctcctccatcagaacccggagagccttaagagaaataacagattt 439087
Oy 690 gtgggacaagcccaaggttcctgtgtgtgcaagcctgaagaaactgtgctcgtgtc 749
Db 439088 gtgggccaagctcccaaggttcctgtgtgtgcaagcctgaagaaactgtgctcgtgtc 439147
Oy 750 cagttcaattcagctgt-----ggctcagaagggtcgaagcc 785
Db 439148 cagctgtctcagctatgtctaaaggaggtgagtaaccaggtttcagaagctc 439207
Oy 786 ccaagccttgcagacttccaagtgtgtgtgagcctgtgtgttccaagaaagtaaaatg 845
Db 439208 ccaaaccttgcagacttcaatgtgtgtgtgagcctgtgtgtacacagaagttcaagaattg 439267
Oy 846 aggttgggaacctccaatcagaattcagaatataatgaaacccctgtagtccagtc 905
Db 439268 aggttgggaacctccaatcagaattcagaatataatgaaatgtagtccagtcagtc 439327
Oy 906 agaagttgtgtgaggt 965
Db 439328 aaaaattgtgtgagaggt 439387
Oy 966 gaaatgt 1025
Db 439388 gaaatgt 439447
Oy 1026 tgaagaagaagtcacacatcctcagactcagaagagtgatccactgagacattgcaagc 1085
Db 439448 taagagagagacactgtctcctcagacccagaaatagatagatccactgagacattgcaagc 439507
Oy 1086 atgtgcttgaanaatccacagacactcagtcgcaagctgttgaaga-aaaaaa-gca 1135
Db 439508 atgtgcttgaanaatccacagacactcagtcgcaagctgttgaagaagcagtcagtcagtc 439567
Oy 1136 gggagtgagctgt 1195
Db 439568 gt 439627
Oy 1196 taacctgtgagcttgagacatgagcttgagacatgagcttgagacatgagcttgagacatgag 1255
Db 439628 taacctgtgagcttgagacatgagcttgagacatgagcttgagacatgagcttgagacatgag 439686
Oy 1256 gcttaagaattgagctgcccacatgagcttgagcttgagcttgagcttgagcttgagcttgag 1314

||||| ||||||| |||||| | ||| | |||||||
Db 439687 gcttaagattgactgcgcgtgagatttaagactgtgtggtccctgaacccttgg 439746
QY 1315 ttctggccatttttccatttgaactgcgcgtatttaaccatgcttacttcattg 1374
Db 439747 ttctggccatttttccatttgaactgcgcgtatttaaccatgcttacttcattg 439806
QY 1375 tatcta 1380
||| |
Db 439807 tatcta 439812

RESULT 10
AAH13678
ID AAH13678 standard; cDNA: 1736 BP.
XX
AC AAH13678;
XX
XX 26-JUN-2001 (first entry)
DT
XX
DE Human cDNA sequence SEQ ID NO:10539.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000BP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 8; SEQ ID 10539; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to a
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX
SQ Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;

Query Match 43.8%; Score 605.2; DB 22; Length 1736;
Best Local Similarity 79.1%; Pred. No. 6.8e-179;
Matches 813; Conservative 0; Mismatches 168; Indels 47; Gaps 6;

QY 392 gaagaattcttaagcagaagaagcattcaagaggtgacttggtgtgttaagagcattc 451
||| |
Db 111 gaagaattcttaagcagaagaagcattcaagaggtgacttggtgtgttaagagcattc 170
QY 452 agttcataagggagcagcacaagagtlcaagaatttgcacctgacaatgtgata 511
||| |
Db 171 agttcataagggagcagcacaagagtlcaagaatttgcacctgacaatgtgata 230
QY 512 aaaaagaataaccatttcttgagggaattcaagctgtgtgcagaatttgcatagt 571
||| |
Db 231 gacaagaataaccatttcttgaggagattcaagcagctgtgtgaaatttctgtaagt 290
QY 572 aattgagagctgaaatgttaactcctaagaacatggggaataatctccttgacatgtcag 631
||| |
Db 291 agcaagagagcctaattgttagtcccaagacatggggaagatgtctccagacatgtcag 350
QY 632 aggtctcacagcagctccatcaataactgtgacctgtgagagccttagagaataatgtttgt 691
||| |
Db 351 agacctcacacagcccccctccatcacagccagag-----aggaataatgtttgt 405
QY 692 gggagaagcccaaggtctccctgtgtgtgtgagcagcctagaagacttggccctgtgtccca 751
||| |
Db 406 gggccctgtgtcagaggtccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 465
QY 752 gtaattc-----agctgtgtgtcagag 775
||| |
Db 466 gctgt 525
QY 776 ggtgcagccccaagagcctgt 835
||| |
Db 526 ggtgcagccccaagagcctgt 585
QY 836 tc-aagaaattgagtttgggaacctccaatcgaatttgggaagatatatgaaccctgt 894
||| |
Db 586 tcaagaatttgaagttgt 642
QY 895 gatgccagagcagaagttgtctgt 954
||| |
Db 643 gatgccagagcagaagttgtctgt 701
QY 955 agtacaaaagggaatgt 1014
||| |
Db 702 agtgtgaagggaatgt 761
QY 1015 tagtagactgtgtgaagaagtcacacatcttcagacactcagaagaggttaatactgtga 1074
||| |
Db 762 tagtagactgtgtgaagaagtcacacatcttcagacactcagaagaggttaatactgtga 821
QY 1075 cagcttgacagatgt 1134
||| |
Db 822 cagcttgacagatgt 881
QY 1135 aggaattgagttgttacacctacaacacgtagtgagagagctgtgacaaagacgttgggaat 1194
||| |
Db 882 cggagggagagctgttacacctacaacacgtagtgagagagctgtgacaaagacgttgggaat 941
QY 1195 ctacctctgtcatgtcatgtacactgtgagcgtgtgagactgtgagtcacaagaagtatttttg 1254
||| |
Db 942 tcaactctgtcatgtcatgtacactgtgagcgtgtgagactgtgagtcacaagaagtatttttg 1001
QY 1255 agcttaagatttgaagcccaactgatttggacttataggggcccgt-acccttc 1313
||| |
Db 1002 agcttaacatttgaagcccaactgatttggacttataggggcccgtacccttc 1061

1D	AAS44506 standard: DNA; 31529 BP.
XX	
AC	AAS44506;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human LEKTI DNA clone C17978SKR_94F21 contig 8, SPINK5 exon 5.
XX	
KW	Human: SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW	serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW	eczema; hayfever; antihistaminic; antiallergic; antiinflammatory;
KW	dermatological; PCR primer; sequencing primer; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	WC200164747-A1.
PD	
PF	07-SEP-2001.
XX	
PR	02-MAR-2001; 2001WO-GB00897.
XX	
PR	02-MAR-2000; 2000GB-0005098.
XX	
PR	03-MAR-2000; 2000GB-0005229.
XX	
PA	(ISIS-) ISIS INNOVATION LTD.
XX	
PI	Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
DR	WPI; 2001-562149/65.
XX	
PT	Determining susceptibility to atopic disease or carrier status of
PT	Netherton's syndrome in humans by identifying variants of or mutations
PT	in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT	inhibitor -
XX	
PS	Disclosure; Page 95-103; 123pp: English.
XX	
CC	Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC	fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC	SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC	(LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC	an atopic disease in a human subject can be detected by screening the
CC	genome for one or more polymorphic variants of SPINK5 gene and/or
CC	expression of a variant LEKTI protein in a tissue. Carrier status of a
CC	subject or development of Netherton's syndrome is diagnosed by screening
CC	for the presence of loss-of-function mutations in the SPINK5 gene. An
CC	expression vector comprising a nucleic acid encoding a serine protease
CC	inhibitor or its functional fragment can be used in screening for
CC	compounds with potential pharmacological activity by determining the
CC	serine protease activity of a protein previously identified as a ligand
CC	of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC	asthma, eczema and hayfever.
XX	
50	Sequence 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;

Query Match	43.3%;	Score 598;	DB 22;	Length 31529;
Best Local Similarity	77.9%;	Pred. No. 6.8e-176;		
Matches 815; Conservative	0;	Mismatches 210;	Indels 21;	Gaps 7;

[illegible]

Db	14328	tcctgagagagaatcacaagccaggttgcataaatcttgcataagtagcaagagagccttgat	14387
Qy	590	aatctcaagaacaaatctgaggaataatctctcttgcacatctgcagagctcttccacagcctcc	649
Db	14388	aatcccaagaacaat-gggaaaaatgtctcccaagagatatacagagactcttctgtagcc	14446
Qy	650	atcaaatcaatgcctgcagagccttagag---aaaatggttcttctgtagcaagccagag	706
Db	14447	ctcccatatacatcgcacagaggtttagaggaataaatgtctctctctagccagagcttagag	14506
Qy	707	tcacctgagctgttgcagcc-----tagaactctgtgtgcctgtctccagttaa	756
Db	14507	tcacctgcgctgtgtgaagccctgcgttgaagacatccatcagctgtctgaaggggcatgac	14566
Qy	757	ttcacgctgtgtcttcacagaggtgtgcacaagcccaagcctctgtagcttccaaagtgtgtga	816
Db	14567	ttggagctatgtcttcgcagagggtgcagaagcctgaagccttgcagcttccacatgtgttga	14626
Qy	817	gctctgtgtgtgcacaagaatcagaatctgaagttcttggaaacctccaatcagattcaagaa	876
Db	14627	gctctgtgtgtgcacaagaatcagaatcagaatctgaagttcttggaaacctccaatcagattca	14683
Qy	877	gatatatggaaaacccctggatgccacggagagaagttctgtcttaggggttgggtccctcatg	936
Db	14684	gatatatggaaaatgtcctgtgatttccagggagaagttctctgttaggggttgggtccctcatg	14743
Qy	937	gagaaactctgcagaaggttagtatacaaaaggaatgttgggtctggagcccccacacagagt	996
Db	14744	gagaaactctgcagagcagtagtacaagaagaaatgttggagtagagagccccacacagagt	14803
Qy	997	ccccagtgggggtcccatctagttagagctctgttggagaagaggttccacatccctccagagctca	1056
Db	14804	ccccactc-gggccacccctctgtgttagagctcttcttggagaagaagcttccatccctccagaacca	14862
Qy	1057	gaagagtagatccacatcgacagctctgcagcatgtgcctcttcaaaaatccacagacacatagtg	1116
Db	14863	gaatagtagatccacaaggttagcttgcacacatagtctgtgagaagaagccacaacacatcagtg	14922
Qy	1117	ccagactgtgaaagcgacgggagttgagatctgttacctacaacaaacgttagtggcagagct	1176
Db	14923	ccagactgtgaaagcgacgttggagggagggcgttaccctgtccaagccacagaaagttagagatt	14982
Qy	1177	gaccaaagacogtvtggagatctactctcttgatctgtcatagactgttagacgttggagatvgagt	1236
Db	14983	gcccacaaacacatvggaaacacacactctgtgatacgaatgacgttgaatagaaatagatt	15042
Qy	1237	caaaagagatcatcttggagccttaagaattgacgtcccccctgagatttggagactaat	1296
Db	15043	caaaatatcatcttggagactttaagaattgactgcgtccctccgtgaatttgcagctgcat	15102
Qy	1297	ggggcccgcta-cccccttcttctggccaatttcttccatcttggaaactgtccgttatccac	1355
Db	15103	ggacatctgtagcccttcttctggccaatatctccacatttggaaagcgttatgtgacc	15162
Qy	1356	aatgcctgtatccctcatgtatgtag 1381	
Db	15163	aatgctgtatcccatgttatctag 15188	

RESULT	13
AAI99255/c	
ID	AAI99255 standard; DNA; 32127 BP.
XX	
AC	AAI99255;
XX	
DT	07-JAN-2002 (first entry)
XX	
DE	Human excretory related polynucleotide SEQ ID NO 1019.
XX	
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW	antiparkinsonian; antischizoid; antianaemic; antiarthritic; cancer;
KW	antiparkinsonian; antischizoid; antianaemic; antiarthritic; cancer;

KM antihypertensive; hepatotropic; cerebroprotective; antiinflammatory;
KM antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KM neurological disease; infection; nephrotropic; gene therapy; vaccine;
KM excretory system; ds.

OS Homo sapiens.

PN MO20015313-A2.

XX 02-AUG-2001.

PD 17-JAN-2001; 2001WO-US01323.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226271.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239933.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465569/50.
XX
PT Isolated nucleic acid molecule encoding excretory system antigen 1s
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Example 2; SEQ ID NO 1019; 574pp + Sequence listing; English.
XX
CC The invention relates to novel excretory system related human
CC polynucleotides (AA198567-AA199503) and the encoded proteins
CC (AA99594-AA99913) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy, especially
CC disorders related to the excretory system. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32127 BP; 8788 A; 6237 C; 6877 G; 10225 T; 0 other;

Query Match 42.9%; Score 592.4; DB 22; Length 32127;
Best Local Similarity 79.1%; Pred. No. 3.9e-174;
Matches 813; Conservative 0; Mismatches 161; Indels 54; Gaps 7;

QY 391 ggaagaatttctaagacaagcaaatcaagagtgactgggtgctgttaagagcatt 450
DB 22042 GGAAGAAATTTCTAAGCAGCAGACATTCAGAGGTGACTGGGTGCTGTTAAAGCAT 21983
QY 451 cagttcataagggagagcagagcgaatgaagttcagaataattgcacctgcacatgcat 510
DB 21982 CAGTTTAAAGGAAACAGACGATAAAGCTTGGAAGTTTGACGCTGACCAATGCAT 21923
QY 511 aaaaaaagaaccatttctgaggggaattcaagctgctcagaatttcattcatg 570
DB 21922 AGAAAGAGAA-----TCTCAGGTGACATTCAGGCACGCTGCAGAGTTTGCAATAG 21671
QY 571 taatgagagctgaatgttaactcctaagacaatgagggaataatctctcctgagacatgca 630
DB 21870 TAAAGAGGGGCCAATGTTATCCCAAGAAATGCGGAAATGTCTCCAGGCGATGCA 21811
QY 631 gaggtcttaagcagagccatcaatactactgctgagagcctggag---aaatggtt 687
DB 21810 GAGATCTTCAATGAGCCCTCCCATCAACAGGCTGACAGGCTGAGGAGAAAGTGGTT 21751
QY 688 ttgtggagacaagggagctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 747
DB 21750 TCAATGGGCGAGGCGCCAGGAGTCCCATGCTGTGTGAAGCTTAGGAGCTTGGCCCTGAT 21691
QY 748 ccagtaattcagc-----tgtgtctcaga 774
DB 21690 CTCAGCGGCTTAGCATGCGCAAGAAAGGGCCCAAGTAGAGCTCAGCATGTGGCTTCAGA 21631
QY 775 ggggtcagaagcccaagccttgagcagcttccaagtgtgtgtgtgtgtgtgtgtgtgtgtgt 834

DB 21630 GGGTCAAGCCCCAAGCCTTGGCAGCTT-CACATGCTGTTAAGCTTCGAAATGCACATTA 21572
QY 835 gtcaagaattgaggttttggaaacctccatcagatttcaagaagatatggaacccctg 894
DB 21571 GTCAAGAACTGGGTTTGGAACTCCACCTAGATTTCAGAAAGTGTGAATGCTGTG 21512
QY 895 gatcccaagcagaagtttctgtgaagggtgggttcctcatcgtgaacaccttcgaaggt 954
DB 21511 GATGCCCGCAGCAGAA-TTTCCTGCATGGGCGAGGCCCTTATGAGAACTCTACTAGGGC 21453
QY 955 agtcaaaaagggaatgt 1014
DB 21452 AGTGTGAAAGGAAATGTGAGGTGCTGTCGCCACACAGAGTTCTACTGAGCACTGCC 21393
QY 1015 tagtaagactgtgaagaagaatccacatcctccagactccagaagaagtaatccactga 1074
DB 21392 TAAATGAGACTGTGAAGAGAGGCGCAGCATCTCCAGATCCAGAAATGTATCATCAGA 21333
QY 1075 cagcttgacgacatgtgcttgaaaaaatccacagacactcagtgcaagcctgtgaagcagc 1134
DB 21332 CAGCTTGACACACGTCCTCGGAAAAAGCTGACAGACATTCAAGACCAAGCCTGTGAAGCAGC 21273
QY 1135 agggatgagctgttaccctacaaaacgtagtggcag-agctgaccaagacgctgggaa 1193
DB 21272 CAGGAGGAAACTGTACCTCGACAGCCACAGGGGTGAACCTGCCAAGACATGGGA 21213
QY 1194 tctaccttgcattgtatcactgtgacgtgagacatgagtgatgaagaagaatatttg 1253
DB 21212 CCCACGGTTTCATCAGCGGTATACCCCGATGGAGACAGGAATCAAGAGATCATTTTG 21153
QY 1254 gacgttaagaattgactgcccacatgtgattcggactatataggggccgtacccttc 1313
DB 21152 GAGCTTTAAGATTTGACGCTCTCACTGATTTCACTT-----GCCGTGACCCCTT 21100
QY 1314 gtttgccaatttttccatttggaaactgcgcatatcccaatgctgacctcatt 1373
DB 21099 TGTTAGCTAAATTTCTCCATTTGGAAATGCGCGTATTTACCAATAGCTGTACCCCAT 21040
QY 1374 gtatgtag 1381
DB 21039 GTATCTAG 21032

RESULT 14
AA163605/c
ID AA163605 standard; DNA; 32127 BP.
XX
AC AA163605;
XX
DT 22-OCF-2001 (first entry)
XX
DE Human kidney related polynucleotide SEQ ID NO 920.
XX
KW Human; kidney antigen; immunosuppressive; antihypertic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; vitucide; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW antineoplastic; anticancer; vulnery; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
OS Homo sapiens.
XX
PN W0200155323-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01343.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.

CC (AA162971-AA163793) and the encoded polypeptides (AA042417-AA042691)
CC collectively known as kidney antigens and the use of such kidney antigens
CC for detecting disorders of the kidney, especially kidney cancer and
CC kidney cancer metastases. The polynucleotides and proteins are also
CC useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. The genes are isolated from a range
CC of human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 32127 BP: 8788 A: 6237 C: 6877 G: 10225 T: 0 other:

Query Match 42.9%; Score 592.4; DB 22; Length 32127;
Best Local Similarity 79.1%; Pred. No. 3.9e-174;
Matches 813; Conservative 0; Mismatches 161; Indels 54; Gaps 7;

QY 391 ggaagaatttctaagcagcaagaagatcagaagtgactgggtgtctgtaaggcatt 450
DB 22042 GGAAGAAATTTCTAAGCAGCAGACGATTCAGAGTACTGGTGCTGTTAAAGGCATT 21983
QY 451 caattcacaaggaggagcagcacaagatgcagaagatttcacccctgacatgtgat 510
DB 21982 CACTTTTAAAGGAAACAGACATTAAGCTTGGGAAGTTTGCAGCCTGACATCAT 21923
QY 511 aaaaaaagaaaaccatttctgaggggaaatccaagctgagcagaagaatttgcatalg 570
DB 21922 AGAAAAGAAAA-----TCTCAGGTGACATTCAGAGCAGCTGCGAAGATTGCTAAG 21871
QY 571 taatgagagcctgaaatgtaattcctcaagaagaatgggaaatatctccgcagcatgta 630
DB 21870 TAATGAGGGGCCAAATGTTAATCCCAAGAAATGGGAAATGTCTCCAGGGCAGTGA 21811
QY 631 gaagcttcacagcagtcacatcaatcactgctgagccctagagag--aaatgctt 687
DB 21810 GAGATCTTCATGGCAGGCCCTCCCATCAGAGCCTGCAGGCTTAGAGGAAAGTGCTT 21751
QY 688 ttgtggagaagcagcaggggtccctgtgtgtgtagcagcctagaagaacttggtccctgt 747
DB 21750 TCATGGGCGAGGCGCAGGGTCCCATGCTGTGTGAGACCTTAGGAGACTTGGTCCCTG 21691
QY 748 ccgaattaatcagc-----tggtctcaaa 774
DB 21690 CTAGGCCGCTTAGCCATGCGCAGAAAGGGCCCACTAGAGCCTCAGCATGTGGCTTCA 21631
QY 775 ggggtgcaagccccaagccttgtagcagcttcagaagtgctgtagcctgtggtgcaaga 834
DB 21630 GGGTGCAAGCCCAAGCCTTGACAGCTT--CACATGGTGTGAGCCTGCAAGTGCACATTA 21572
QY 835 gtcaagaattgaggttggaacctccaatcagaatttcagaagatatagaaaccctgt 894
DB 21571 GTCAGAACTGGGGTTGGGAACCTCCACTGATTTTCANAGATGATGGAATGCTCTG 21512
QY 895 gatgcccagcagaagtttgctgtaggggtggtgctcttgtagaaacctgtcaagggt 954
DB 21511 GATGCCAGCGCAAA--TTTGTCTGCTAGGGCAGGCCCTTATGGGAACCTCTACTGTCG 21453
QY 955 agtacaagaaggaaatgtgtgtgtagagcccccacacagagtcgccagtggagctcac 1014
DB 21452 AGTGTGAAGAAGAAATGTAGGTGCTGCCCCACACAGAGTTCTTACTGTGACATGTGC 21393

QY 1015 tagtagagctgtgaaagaagtcaccacatcctccagactccagaaggtagatcactgta 1074
DB 21392 TATGAGCGCTGGAAGAAAGAGGACGACGATCTCCAGATCCCAAGATGTAGATCCACTGA 21333
QY 1075 cagcttgacagctgtgcttgaaaatccacagacactcagctgtgcagcctgtaaagcgc 1134
DB 21332 CACCTTGCACCCAGCTGGTGAAGAAAGCTGACATTCAGACACACCTGTGAAAGCAGC 21273
QY 1135 aggaagtagctgtgacccctacaacacccgtagtgcaag--agctgacaaagcctgggaa 1193
DB 21272 CAGGAGGAAATGTATACCTCGACAGCAGCGGTGAGAGCTGCCCAAGACATGGGAA 21213
QY 1194 tctaccttgcatgtgcatgacctgagcgtgagcatgtagacatggaagaagatcattg 1253
DB 21212 CCACCGCTTGATCAGGCTGACCGGATCGGAGACACGGAATCAAGAGATCAATTTTG 21153
QY 1254 gagcttaagaattgacacgtgcccacatgatttggaattataggggcccgtaccctt 1313
DB 21152 GACCTTTAAGATTGACATGCTCCTCAGATTTTCACACTT-----GCCTGTAGCCCTT 21100
QY 1314 gtttgccaatttttccatttgtaactggaactgcgtratttaccacatgcctgtaacctat 1373
DB 21099 TGTTAGCTAATTTCTCCATTTGGAAATGGCGGATTTACCAATGCTGTACCCCAAT 21040
QY 1374 gtagttag 1381
DB 21039 GATCTTAG 21032

RESULT 15
AAK86832/C
ID AAK86832 standard; DNA: 72215 BP.
XX
AC AAK86832;
XX
DT 07-NOV-2001 (first entry)
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:41644.
XX
KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

Query Match 42.9%; Score 592.4; DB 22; Length 72215;
Best Local Similarity 79.1%; Pred. No. 6.4e-174;
Matches 813; Conservative 0; Mismatches 161; Indels 54; Gaps 7;

OY 391 ggaagaatttctaagcagcaagcattcaagaagtgtactgtggtcgtttaaagcatt 450
|||||
Db 22042 GGAAGAAATTTCTTAAGCAGCAGCATTCMAAGAGTGGTGTCTGTTAAAGCATT 21983
OY 451 cagttcataaaggagcagacataaaggttcagaaaattgcacccctgacatgtgat 510
|||||
Db 21982 CAGTTTAAAGAAAGCAAGCAGCATATAAAGCTTGGGAAGTTGCAGCCTGACATGCAAT 21923
OY 511 aaaaaaagaaacccatttctcgtagggaaalticaagctgtcgtcagaaaatttgcatatg 570
|||||
Db 21922 AGAAAGAAAA-----TCTCAGGTGACATTCAAGCAGCTGCAGAGTTTGCATAG 21871
OY 571 taatagagagctgaatgttaatccccaagacaaatgaggaaaataatccttgacatgtca 630
|||||
Db 21870 TAAATGAGGGGCCAAATGTTAAATCCCAAGAAATGGGAAATGTCTCCAGGGCATGTCA 21811
OY 631 gaggtcttcacagcagtcacatcaactgcctgagcctgagagcctagagag---aaatggtt 687
|||||
Db 21810 GAGATCTTCATGAGCCGCCCTCCATCAAGGCTGAGGCTTAGAGSAAATAGTGTT 21751
OY 688 ttgttgagacagcccaaggttccctgtgtgtgtcagcctagagacttgggtccctgtgt 747
|||||
Db 21750 TCATGGGCGAGGCCAGGCTCCCATGCTGTGTGAAGCCTAGGACTTGTGCTCCCTGCAT 21691
OY 748 cccaagtaattcagc-----tggtgttcaga 774
|||||
Db 21690 CTCAGCCCGCTTAGCCATGCGCAAGAGGGCCACGTAGAGCTCAGCATGTGCTTCAGA 21631
OY 775 ggggtcagaagcccaagccttggcagcttcacaagtgtgtgtgagcctgtgtgtcagaagaa 834
|||||
Db 21630 GGGTCCAGCCCAAGCCTTGCGACAGCTT-CACATGCTGTGAGCCTGCAAGTGCACATTA 21572
OY 835 gtcaagaattgaggttgggaaacctccaatcagatttcaagaagatatatggaacccctg 894
|||||
Db 21571 GTCAAGAACTGGGGTTTGGGAACCTCCACCTAGATTTCAGAAATGTAGAAATGCCCTG 21512
OY 895 gatgccagcagcagaagttgtcgttaggggtgtgtccatgtgagaacacctgtcagaaggt 954
|||||
Db 21511 GATGCCCAGCGAGAA-TTTGCTGCGATGGGCGAGGSCCTTATGAGAACCTCTACTAGGCG 21453
OY 955 agtacaaaagggaatgttgggtggagccccacacagagtcgccagtggggtccatc 1014
|||||
Db 21452 AGTGTGAAGAAAGGAATGTAGAGTGGTGGTCCCCACACAGAGTTCTACTGAGCACCTGCC 21393
OY 1015 tagtagagctgtgagaagaagttccacacatccctccagactccagaaggttagatccactga 1074
|||||
Db 21392 TAAATGAGCTGTGAGAGAGAGGCCAGCATCTCCAGATCCCGAATGGTAGATCCACTGA 21333
OY 1075 cagcttgacagcatgtgccttgaanaaaaccacagacactcagtgccagcctgtgaagcagc 1134
|||||
Db 21332 CAGCTTGCAACGCGTGCCTGGAAAGCTGCAACATTTCAAGACAGCTGTGAAGACAGC 21273
OY 1135 agggatgagttcgtacctacaanaaocgltagtgag-agctgaaccaagaccgtlgyaa 1193
|||||
Db 21272 CAGGAGGGAAACTGTACCTCGAGACACAGGGGTAGAAGCTGCCAAGACCATGGGA 21213
OY 1194 tctaccttctgattgtcatgacctggagcgtgagacatgagatcaaaagagatcatttg 1253
|||||
Db 21212 CCCACCGCTTGCATCAGGTGACCCGGATGCGAGACAGGATCAAGAGATCATTTTG 21153
OY 1254 gagcttiagaatttgacccccactgatttgcgaattatataatggtgcccgtacccctt 1313
|||||
Db 21152 GAGCTTTAAGATTGACTGCTCACTGAGATTTCACACTT-----GCCGTGAGCCCTT 21100
OY 1314 gttttggccaatttttccatttggaaactgcglatattaccaatgacctgtaccccat 1373
|||||
Db 21099 TGTTAGCTAATTCTCCATTGGAATGGCGGTATTATCCCAATGCTGTACCCCATT 21040

OY 1374 gtagtag 1381
|||||
Db 21039 GTATCTAG 21032
Search completed: June 30, 2002, 14:44:24
Job time: 4265 sec

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TREATMENT AND DIAGNOSIS OF BREAST CANCER
 NUMBER OF SEQUENCES: 292
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed IP Law Group
 STREET: 701 Fifth Avenue, Suite 6300
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/991,789A
 FILING DATE: 11-Dec-1997
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Potter, Jane E. R.
 REGISTRATION NUMBER: 33,332
 REFERENCE/DOCKET NUMBER: 201021.419C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 620 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 29:
 GS-08-991-789A-29

Query Match	17.9%	Score 247.4	DB 4	Length 620:
Best Local Similarity	68.6%	Pred. No. 1.7e-69:		
Matches 424:	Conservative 0:	Mismatches 177:	Indels 17:	Gaps 6
QY 772	agaaggtgcaagccccaagccttgcagcttcccaagtgtgtcttgagccctgtgtgtcaaa	831		
Db 2	AGAGGTCACAGCCCAAGCCTTGACCACTTCATAGGGTGTGCACAGCCTGTGGTGCACA	61		
QY 832	gaatgcaagaaattgaggttctgggaacctcccaatcagatttcgaagaatatgaa-----	886		
Db 62	GAAGTCAAAAAATTGAGTTTGGATTCCTCAGCCTGATTTTCAGAGAGTATTAAGAACAAC	121		
QY 887	-aaacccttgatgcccgcgcgaagaattgtcgtgtagggtgtgggtccctatagagaacct	944		
Db 122	CTAACACCTTACATATATTTTCAGCAAAAAGTTTACTACAGGGATGAAGCTTTTCACGAAAACTT	181		
QY 945	ctgcaaggttagtacaaaggaatgtctgtgtgtagagcccccacacagatgccccagtg	1004		
Db 182	CTACTAGGAAGTGTACAGAGAGAAATGTGGGTTTGGAGGCCCCCAACAGAAATCCCTCTA	241		
QY 1005	ggagctccatctagtagagctgtgagaagaagtccacacacctccacagatcccaagagta	1064		
Db 242	GAACACTGCTTAATGAAACTGTGAGAGAAGTGCCCACTGTCATCCAGACACAGAAATGATTA	301		
QY 1065	gatccatctgacagctctgcagcat-gtgccttgaaaaaatccacagacactcagtgcagcc-	1122		
Db 302	GACCACCCAAAAACCTATTAGCCATATTTGGCTTTAAAAACCTACAGACACTCAATGCGACGCC	361		
QY 1123	-ttgtgaagaagcaggagatgagctcttaccctcaaaaacccgtagtgtgcaagctgacc	1180		
Db 362	CATGAAAAAAAACGAGAGAGAGACTGTGTCCTTCAAAATGCGACCGGAGCAACACTGCC	421		
QY 1181	aagaacctgggaatcttaccctctgtcatctgtcactgagcctgagacg-gaagacaggaattca	1239		
Db 422	CAGGCAATGGAAGACGACACTCTTATATCAATGTAGACTGTGATGCTTGAGACATGGAAATCCN	481		
QY 1240	aagaagatcatcttggagctttaagat-----tgactgcccaactgattctgacttata	1295		

[illegible]

```

RESULT 4
US-09-062-451-29
: Sequence 29, Application US/09062451
: Patent No. 6344550
: GENERAL INFORMATION:
: APPLICANT: Fridakis, Tony N.
: APPLICANT: Smith, John M.
: APPLICANT: Reed, Steven G.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
: NUMBER OF SEQUENCES: 297
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/062,451
: FILING DATE: 04-Apr-1997
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: MAKI, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.419C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 29:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 620 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-09-062-451-29

```

Query Match	17.9%	Score 247.4	DB 4	Length 620
Best Local Similarity	68.6%	Pred. No. 1,76-69		
Matches 424	Conservative 0	Mismatches 177	Indels 17	Gaps 6
OY	772	agaaggtcaagagcccaagccttgcagcttccaaagtgtgttgacctgtgtgtgcaaa	831	
Db	2	AGAGGGTCCAGCCCAAGCCTTGACACCTTCATTAAGGGTGTCAAGCCGTGGGTGCACA	61	
OY	832	gaagtcagaagttagtggtttggaacctccaaatcagattccagaagatatactgaa-----	886	
Db	62	GAAGTCAAAATTTGAGTTTGGATCTCCAGCCTTGATTTGAGAGATATTAAGAACAAC	121	
OY	887	--aaccttgatgcccacggcagaagtttcgtgtgggttggtggtccatattgaaacct	944	
Db	122	CTAATACCTTAGATTATTCAGACAAAAGTTTACTACAGGGGTGAAGCTTTACCGAAAACCT	181	
OY	945	ctgcgaaggtagttacaaaaggaaatgtttggttggagagcccccacacagatgcccaagt	1004	

```
Db 182 CTACTGGAAGTACAGAGAAATGTGGTTTGGAGCCGCCCAACACAGATCCCTCTA 241
Qy 1005 gggctccatctagtagagctgtgagaagaagtcacacatccctccagactccagaaggtga 1064
Db 242 GAACACTGCTTAATGAACTGTGAGAAAGATGGCCACTGTGATCCAGACACAGAAATGATA 301
Qy 1065 gatccactgacagctgtgcaagat-gtgcctgaaataacccacacactcaagtgccagcc- 1122
Db 302 GACCCACCAAAACTTATGCTATATGCTATATAAACCTACACACTCAATGACCAAGCC 361
Qy 1123 --tgtaagaagcagcaggatgagatctgtacatacaaacgtaatgacagactgac 1180
Db 362 CATTGAAAAAAACCTGAGAGAAAGACTGTGCTACATATCCACCGAGACAGAACTGCC 421
Qy 1181 aagacccgtggaatctaccctctgcatgtcatgacctggaag- tgaacatgagatcaa 1239
Db 422 CAGGCGATGGAAGCACACACTCTTATATCAATGTGACTGTGATGTGAGACATGAAATCCN 481
Qy 1240 aagaatcatctttagagctttaagatt---tgactgcccactggaatttggaattata 1295
Db 482 AAGCAATCTTTTAAACCTTCCACGGTTMAATGACTGCCCTATTATTAATTCNAACCTTANA 541
Qy 1296 t-gggagccgtaacccttctgtttgagccaatttttaccatttggaaactgcgtattacc 1354
Db 542 TCCNGGCGCTGTGACCTCTTTGCTTTGGCCATTCGCCCTTTTGGATGGCTTTTTC 601
Qy 1355 caatgcctgatacccat 1372
Db 602 CCATGCCGTGCCCTCTT 619

RESULT 5
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)---(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16
```

Query Match 17.4%; Score 240.2; DB 3; Length 152331;

Best Local Similarity 59.4%; Pred. No. 1e-65; Matches 320; Indels 71; Gaps 7;

```
Matches 572; Conservative 0; Mismatches 320; Indels 71; Gaps 7;
Qy 445 ggcattcagttcatcaagagagcagacagagtcagaaatttgcacctgacaa 504
Db 71252 ggaacctatggttaaatggaagtagagttcaaaatttggaaatttgcagccggtc 71311
Qy 505 tgtgataaaaaaagaacccatcttctgaggggaaattcaacgtgctgcagaaatttg 564
Db 71312 tctgagcagaagaatccaagtgcgtgcagagcatcatgtctag---agagattag 71367
Qy 565 catatgtatgagagcgtgaatgttaattcccaagaacatggggaataataatctctgac 624
```

```
Db 71368 catgtctaaagggagccaaagtctctaataatcaagaacatgtttaaagccttgaggc 71427
Qy 625 atgtcagaagctctcaagaagagtcacatcaatcaactgacctgagagcctag-----ga 677
Db 71428 attcagaagctatagtagcagagccctccatcacagtgagagagtttggctgacatag 71487
Qy 678 gaaatgttttctggagcagagcccgaggttccctgtgctgtgtgtgtagctctagaacttg 737
Db 71488 ccagaagtttctttagggccanngcagggccacatgcatgacagcttvgaacactg 71547
Qy 738 tgcctgtgtcccgcttaa----- 756
Db 71548 cgcgccgatattccagccacactctgctgtctccaccccttgctcaaacggccaagatag 71607
Qy 757 ---ttcaactgtgtgtcttcagaagggtgcaagcccaagccttvgacacttccaagtgt 813
Db 71608 agcttggacacactctcccgagggcacaagaactaagccttggttccatgtgtgt 71667
Qy 814 tgaactgtgtgtcagaagaagtcagaat--tgaagtttggaaacctccatcaagatt 871
Db 71668 taagcctgcaggtgccaagaatgcaagatgagggagcttggcacctccactaaatt 71727
Qy 872 cagaagataatggaacccctgagatgcccagcagaagtttgcgtaggggtgtgtcc 931
Db 71728 cagagagatgttcaagaacacttaggttccaggcagaagcatgatacaggggcagagcc 71787
Qy 932 tcaatgagaacctctgcagaaggtagtacaagaaggaaatgttgggtggagcccccac 991
Db 71788 ttgcagagaaacctctactactagggcaatgccaagaataatgttgggttggatctccac 71847
Qy 992 aagatcccccagttgggtccatctctagtagagctgtgagaagaagtcacacatccctcaga 1051
Db 71848 atgtgccacactggggacactcactgtgtatctgtggaaatggggctgtgcctccaga 71907
Qy 1052 ctccagaaggttagatccactgacagacttgcagcatgtcctgaaataatccacagacat 1111
Db 71908 cccagaatgttatatgtaactgtgcagctgtgcacccctggacccggaaaagctgcagacat 71967
Qy 1112 cagtgcagcctgtgaaagcagcagggatgtgaatctgtacccctcaaaaaacgtgtgca 1171
Db 71968 caactccaacccaatgagatcagccacatggtctac--tccagggaaagcccaagagca 72025
Qy 1172 gagctgaccaaagacccgtgggaatctacccttgcattgtcattgacactggaagtgaagat 1231
Db 72026 gggcgttcaagccttggagcctacccctgt-----aaccaacttgcagagcat 72076
Qy 1232 ggaatcaaaagatcatctttagagctttaagaattgactgcccacactggaattcgact 1291
Db 72077 ggaatcaa---agattatgttgcagcctttagagcttaagctaaatgtttccctgcaattcag 72133
Qy 1292 tatatggggccgtaacccttgttttggccaatttttccatttggaaactgcglatat 1351
Db 72134 cttgtgtggagccgttgccttttttttttttttttttttttttttttttttttttttt 72193
Qy 1352 acc 1354
Db 72194 acc 72196
```

RESULT 6

US-08-687-080-59

; Sequence 59, Application US/08687080

; Patent No. 5965427

; GENERAL INFORMATION:

; APPLICANT: Gregory Dolganov

; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

; NUMBER OF SEQUENCES: 175

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dehlinger & Associates

; STREET: 350 Cambridge Avenue, Suite 250

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

QY 599 gaaacatgggaaacatcttcctggacatgctcagagcttctaacgacgtccaatcaatca 658
Db 1222 RRR 1163
QY 659 ctgcctgagagctcctagagaaatggttcttgagacagcccaag 706
Db 1162 RRR 1115

RESULT 8

```

US-08-463-911-6
: Sequence 6, Application US/08463911
: Patent No. 5868330
:
: GENERAL INFORMATION:
: APPLICANT: Scherer, Philipp E.
: APPLICANT: Lodish, Harvey F.
: TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
: TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millita Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02173
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/463,911
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WM195-05
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 861-6240
: TELEFAX: (617) 861-9540
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1313 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 73..804
: US-08-463-911-6

```

Query Match	3.58;	Score 48.4;	DB 2;	Length 1313;
Best Local Similarity	48.88;	Pred. No. 2.1e-05;		
Matches 164; Conservative	0;	Mismatches 166;	Indels 6;	Gaps 1;

Dd	455	TGCCCATTCGCTTTACCAAGATCTTTCACATATAGCAAAAACACACTGTATGGCTCCACTG	514
OY	282	gaagctcttgccttgcaagagtgctcgtggaaattactactcacaagcttgaatgttgaagctgac	341
Dd	515	GTAATTTCCACCTGCAACATCTCTGGCTGTATCTTGTGCTTACCCACATCAAGTCTATA	574
OY	342	atgcgaagtgcaataatttgctcaatgaggaagcaaa	377
Dd	575	TGAAGGATGTGAAGGTGACGCTCTTTCAGAAAGGACA	610

RESULT 9

```

US-09-140-804-9
: Sequence 9, Application US/09140804
: Patent No. 6197930
:
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Humes, Jacqueline M.
: TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
: FILE REFERENCE: 97-49
:
: CURRENT APPLICATION NUMBER: US/09/140,804
: CURRENT FILING DATE: 1998-08-26
: EARLIER APPLICATION NUMBER: 60/056,983
: EARLIER FILING DATE: 1997-08-26
:
: NUMBER OF SEQ. ID NOS.: 47
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ. ID NO. 9
:
: LENGTH: 4517
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
US-09-140-804-9

```

Query Match	3.5%;	Score 48.4;	DB 4;	Length 4517;
Best Local Similarity	48.8%;	Pred. No. 4.4e-05;		
Matches 164; Conservative	0;	Mismatches 166;	Indels 6;	Gaps 1

RESULT 10

US-09-188-930-217
; Sequence 217, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells

;; TITLE OF INVENTION: and Methods For Their Use
;; FILE REFERENCE: 11000.1011c1
;; CURRENT APPLICATION NUMBER: US/09/188.930A
;; CURRENT FILING DATE: 1998-11-09
;; NUMBER OF SEQ ID NOS: 348
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 217
;; LENGTH: 1107
;; TYPE: DNA
;; ORGANISM: Rat
US-09-188-930-217

Query Match 3.1%; Score 42.4; DB 3; Length 1107;
Best Local Similarity 50.5%; Pred. No. 0.0016;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 169 agaaagcagccttactgtaagctcagtgaaactctctctccttcaagccat 228
DB 614 agccaagtcgacctcttcggtcggtacccaagagttaccacgtgagcactgccat 673
QY 229 catctccaaagggctcgtacaaatgccagaaggaattaaaggaagccatggaagctt 288
DB 674 caagttgacaagatctcgtatgaatgagggagccactacaatgcatcagtgccaagt 733
QY 288 tgcctcgaaggtgcttggaattactactcagcttgatgttgagctgcatcattgcaa 348
DB 734 cgtcgcagcgtgcaggaatctacttactctatgacatgacatgacgccaacaaca 793
QY 349 ggtgaatttggtcctaagaa 372
DB 794 cctggccatcgccctagtcacaa 817

RESULT 11

US-08-463-911-1
; Sequence 1, Application US/08463911
; Patent No. 5869330

;; GENERAL INFORMATION:
;; APPLICANT: Scherer, Philipp E.
;; APPLICANT: Lodish, Harvey F.
;; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
;; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Millita Drive
;; CITY: Lexington
;; STATE: Massachusetts
;; COUNTRY: USA
;; ZIP: 02173

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/463.911
;; FILING DATE:

;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: WH195-05
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 861-6240
;; TELEFAX: (617) 861-9540
;; INFORMATION FOR SEQ ID NO: 1:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1276 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 46..786
US-08-463-911-1

Query Match 2.9%; Score 40; DB 2; Length 1276;
Best Local Similarity 49.1%; Pred. No. 0.01;
Matches 106; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 162 ctgatatagaaggtcagccttactcgtgaagctcagtgaaaactctctcttca 221
DB 377 cttatattgattcgtcctcaggtcaggtcggcggaagaccgcctcactgttcccaatg 436
QY 222 agccatcatcttcaagaggtcctgtacaatgccagaaggaattaaaggaagccatg 281
DB 437 taccatttgccttttacttaagattcttacaacacagatcattttgacggagcagc 496
QY 282 gactcttcttcagaggtgctcgtgaattactactcagccttgatgttgagctgcatc 341
DB 497 gcaagtttcactgcacacattccggactcttacttcttaccacatcacggtgtaca 556
QY 342 attgcaagtgtaatttttgctaatgaggaagaa 377
DB 557 tgaaagatgtgaaggtgacgttcttcaagaaagaca 592

RESULT 12

US-08-543-246B-8
; Sequence 8, Application US/08543246B
; Patent No. 6262244

;; GENERAL INFORMATION:
;; APPLICANT:
;; TITLE OF INVENTION: DNA and amino acid sequence specific for
;; TITLE OF INVENTION: natural killer cells
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Michael W. Glynn
;; ADDRESSEE: No. 6262244artis Corporation
;; STREET: 564 Morris Avenue
;; CITY: Summit,
;; STATE: NJ
;; COUNTRY: US
;; ZIP: 07901-1027

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/543.246B
;; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/676,663
;; FILING DATE: 28-MAR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/02469
;; FILING DATE: 27-MAR-1992

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/122,514
;; FILING DATE: 24-SEP-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kasseno, Melvyn M.
;; REGISTRATION NUMBER: 26,389
;; REFERENCE/DOCKET NUMBER: 118-7704/PCT/CONT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 908-522-6927
;; TELEFAX: 908-522-6955

;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1755 base pairs
;; TYPE: nucleic acid

```
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 339..986
NAME/KEY: mat_peptide
LOCATION: 339..986
US-08-543-246B-8
```

```
Query Match
Best Local Similarity 62.2%; Score 38.8; DB 4; Length 1755;
Matches 61; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 1284 ttgcgaactataggagccgtaacccttgcttggccaattttccatttggaactg 1343
Db 1564 TTTGTACTTGTGAGTCCATGACTGTTCTCTCCCTCTTCTCTTGGAAATAG 1623
QY 1344 ccgattaccgaatcgctgaacctgattgtag 1381
Db 1624 TAATATCCATCCTATGTTGTGCTCCACTATGTATTTTG 1661
```

```
RESULT 13
US-09-140-804-1
; Sequence 1, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (198)...(926)
US-09-140-804-1
```

```
Query Match
Best Local Similarity 51.5%; Score 37.4; DB 4; Length 1347;
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```

```
QY 206 ctctctctcttcaagccatcatcttcacagggctctgtacaattgccagagagat 265
Db 543 cctccgcgcgtcgaagcccttcgacccgctgctgtgtgaacagcagagacat 602
QY 266 ttaagagagcagtgagcttctgtcagaggtgctggaatactactaccagctt 325
Db 603 taagaagccgtaacaggaagttaacctgagtggtggttctactactcgcgcgc 662
QY 326 gatcttgagctgacatcatgcaagtgataattgctaatgagaa 372
Db 663 catgcacccgtctacccggcagcagctgagttgattctgtgaaga 709
```

```
RESULT 14
US-09-140-804-10
; Sequence 10, Application US/09140804
; Patent No. 6197930
```

```
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Humes, Jacqueline M.
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
FILE REFERENCE: 97-49
CURRENT APPLICATION NUMBER: US/09/140,804
CURRENT FILING DATE: 1998-08-26
EARLIER APPLICATION NUMBER: 60/056,983
EARLIER FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 729
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate nucleotide sequence encoding the zslg39
US-09-140-804-10
```

```
Query Match
Best Local Similarity 27.9%; Score 36.2; DB 4; Length 729;
Matches 86; Conservative 33; Mismatches 189; Indels 0; Gaps 0;
```

```
QY 65 gcacacccagagcccccagaagaagtgaggctctctgtgcaacaggtttaccacaatat 124
Db 205 ggmngcngngytnccngncmngngaycngncmngngaycngncmngngaycngncn 264
QY 125 aaagagaataagtgatgacaataatgccctctcgtatgataagagagacccctt 184
Db 265 gcngncnccngncngcngngngartgysngtncnccmngnsgntlywsngcn 324
QY 185 actgtgaagctcagtgagaaactctctctcttcacagccatcatcttcacaggggtc 244
Db 325 aarmgnswngartgngngtncncnccmngngaycngncncttctgagymngt 384
QY 245 cgtacaatgccagagagatttaagagagcagtgagcttctgtcaggggtgct 304
Db 385 yngtlnaaygarcarngcaytayagcngtnacngnaartlyacnlygarngcn 444
QY 305 gggattactactcaagcttgatgagctgacatgcatgcaagtgtaatatgtgcta 364
Db 445 gngntataytatygcngtncaycngtntaymngcnmngntncartlygayt 504
QY 365 atgagaa 372
Db 505 gtnaaraa 512
```

```
RESULT 15
US-08-627-151A-15
; Sequence 15, Application US/08627151A
; Patent No. 5866341
; GENERAL INFORMATION:
; APPLICANT: SPINELLA, Dominic
; APPLICANT: BECHERER, Kathleen
; APPLICANT: BROWN, Steven
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: SCREENING DRUG LIBRARIES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Gen-Probe Incorporated
; STREET: 10210 Genetic Center Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```



```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/627,151A
: FILING DATE: 03-APR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fisher, Carlos A
: REGISTRATION NUMBER: 36,510
: REFERENCE/DOCKET NUMBER: CB1016
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-410-8926
: TELEFAX: 619-410-8928
:
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1074 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-627-151A-15

```

```

Query Match      2.6%; Score 35.8; DB 2; Length 1074;
Best Local Similarity 54.1%; Pred. No. 0.2;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 11 ataccctgtcttaatacgcagctcatgtgacatgtagaagttgctgacgtccagcacac 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 768 ATATCGGGGTGAACGGTCAAAAGACATTCACAAACATGATGTCACAGACCTCCAGCATCA 827

Qy 71 cccaggccccccagaaagtggtggcctcctggtgacaccaggtttacacatatatacagga 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 828 CTGTGTCATCCACGACGCGCTGAGCGGCGCTGAGGCACGCTGTCAGCTTCGTGCCCAGGA 887

Qy 131 gaataagtgaaatg 145
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 888 GGAGTTGGGGCAAGG 902

```

Search completed: June 30, 2002, 14:37:29
Job time: 6695 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 11:48:13 ; Search time 1615.67 Seconds
(without alignments)
11536.573 Million cell updates/sec

Title: US-09-997-610-1

Perfect score: 1381
Sequence: 1 gatagtgcatcacctgtct.....tgacctccatgtatgtag 1381

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrc:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	677.8	49.1	2615	10 BE420422	BE420422 32-393 hu
2	615.8	44.6	2009	11 BC004496	BC004496 Homo sapi
3	576.8	41.8	1514	10 BE512633	BE512633 32-1514 h
4	483	35.0	1036	10 BM471183	BM471183 AGENCOURT
5	479.2	34.7	870	10 BI488505	BI488505 603021014
6	477.8	34.6	898	10 BM457166	BM457166 AGENCOURT
7	462.6	33.5	1050	10 BM472108	BM472108 AGENCOURT
8	430.4	31.2	717	12 AG086951	AG086951 Pan trogl
9	419.8	30.4	676	12 AG064424	AG064424 Pan trogl
10	411.4	29.8	654	12 AG112666	AG112666 Pan trogl
11	411	29.8	721	10 BE728616	BE728616 601563827
12	407.8	29.3	935	10 BG681769	BG681769 602627877
13	405.4	29.4	760	10 BG281182	BG281182 602402107
14	404.6	29.3	736	10 BI754555	BI754555 603023384
15	404.2	29.3	691	12 AG062447	AG062447 Pan trogl
16	403.8	29.2	900	10 BM449699	BM449699 AGENCOURT
17	402.2	29.1	1182	10 BM451317	BM451317 AGENCOURT

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
	401.6	401.2	401.2	398.4	397.6	394.4	394.2	394.2	393.8	392.4	390.8	389.4	389.2	388.8	387.8	386.6	383.4	381	380.6	380.2	379.8	377	376.6	373.6	373.6	373.2	373.2	369.6
	29.1	29.1	29.1	28.8	28.8	28.6	28.5	28.5	28.5	28.4	28.3	28.2	28.2	28.2	28.1	28.0	27.8	27.6	27.6	27.5	27.5	27.3	27.3	27.1	27.1	27.0	27.0	26.8
	729	701	704	702	697	699	833	833	877	659	639	676	623	885	704	671	671	742	660	686	940	625	739	760	858	9	675	746
	10 BM45873	12 A058873	12 A058828	12 AG116938	12 AG018926	9 AL602279	9 A0140358	10 BI916691	12 AG097258	12 AG092858	12 AG115932	12 AG125093	12 A0037711	10 BG761979	12 A0020199	12 AG001229	12 AG045131	12 A0748577	12 AG057341	12 AG119684	12 BG679573	12 AG070563	12 AG001404	12 AG001225	12 A0122114	12 B94403	12 AG058688	12 AG001408
	BM45873 603624875	A058873 C17B1-E1-	A058828 C17B1-E1-	AG116938 Pan trogl	AG018926 Homo sapi	AL602279 DKFP686B	A0140358 A0140358	BI916691 603178655	AG097258 Pan trogl	AG092858 Pan trogl	AG115932 Pan trogl	AG125093 Pan trogl	A0037711 C17-HSP-2	BG761979 602718930	A0020199 C17-HSP-2	AG001229 Homo sapi	AG045131 Pan trogl	A0748577 HS-5539_A	AG057341 Pan trogl	AG119684 Pan trogl	BG679573 602627616	AG070563 Pan trogl	AG001404 Homo sapi	AG001225 Homo sapi	A0122114 A0122114	B94403 C17-HSP-217	AG058688 Pan trogl	AG001408 Homo sapi

ALIGNMENTS

RESULT 1
BE420422 LOCUS 1
DEFINITION 32-393 human bone marrow CDNA library Homo sapiens CDNA, mRNA
ACCESSION BE420422
VERSION BE420422.1 GI:16041640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Lu,X., Cui,L., and Li,Y.
TITLE DDRT-PCR from B cell
JOURNAL Unpublished (2000)
CONTACT: xingwu lu, liangxian cui, yonghai li
Department of Biochemistry
Institute of Basic Medical Science, Peking Union Medical College
Dongdan Sanliao 5, Beijing, P.R.C, 100005
Tel: 86-010-65296951
Email: luxingwu@263.net
Full-length and coding sequence.
FEATURES
source location/Qualifiers
1..2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="human bone marrow CDNA library"
/tissue_type="bone marrow"
BASE COUNT 705 a 531 c 641 g 738 t
ORIGIN

Query Match 49.1%; Score 677.8; DB 10; Length 2615;
Best Local Similarity 80.5%; Pred. No. 1.9e-182;
Matches 859; Conservative 0; Mismatches 167; Indels 41; Gaps 4;
Cy 352 gaatttcggcgaatgaagaaatattgctataaaggaagaatttcgaagcaga 411
Db 493 GAACTTGAGCTTGAGAGAGATCAATTAGGTAATCGGCAGAGAAATTTCTAACGACGA 552

OY	412	aagca	tcaaga	gagtg	acttgg	gtgc	gtgt	ltaagg	catc	agtl	tcat	aagg	gagc	aga	471																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
Db	553	AAGCA	TTCAAG	GGGTG	ACTTGG	TGGT	TGGT	CTCT	GTAA	AGCAT	TTACT	TTTATA	AGGA	GACAG	612																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
OY	472	gataa	gggtl	cca	aaatt	tg	cacc	ccg	aca	at	tg	taaa	aaaga	aaac	ccattc	531																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
Db	613	GCAT	AAAG	TTTGG	AAAA	TTTTC	AAC	CTG	AC	AT	CA	TAG	AAAA	GA	AAATCC	ATATTC	672																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
OY	532	tga	ggg	aaat	cca	agct	ctg	ctc	gag	aaat	lt	gc	a	at	l	gc	591																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Db	673	TGAG	GA	GA	AAAT	CA	AG	CTG	CTG	CA	GA	AA	TTTGC	A	TAT	GT	TA	732																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
OY	592	tcct	caaga	caat	tg	gg	aaata	at	ct	ct	gc	at	gt	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca	ag	gt	ct	ca

DEFINITION	Homo sapiens, Similar to hypothetical protein FLJ14058, clone IMAGE:3831313, mRNA.
ACCESSION	BC004496
VERSION	BC004496.1 GI:14709139
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 2009)
JOURNAL	Strausberg, R. Direct Submission Submitted (12-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC/DCTD/DRP CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbioology.org Contact: amadan@systemsbiology.org Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Ketterman and Anuradha Madan
<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov</p> <p>Series: IRAL Plate: 14 Row: d Column: 7</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis</p> <p>This clone has the following problem: frame shifted.</p>	
FEATURES	Location/Qualifiers
source	1..2009
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3831313"
	/tissue_type="Skin, melanotic melanoma."
	/clone_lib="NIH_MGC_20"
	/lab_host="DH10B-R"
	/note="Vector: pOTB7"
BASE COUNT	556 a 429 c 513 g 511 t
ORIGIN	
<p>Query Match 44.6%; Score 615.8; DB 11; Length 2009; Best Local Similarity 78.1%; Pred. No. 9.6e-165; Matches 783; Conservative 0; Mismatches 207; Indels 13; Gaps 3;</p>	
QY	392 gaagaattctcaagcagcaagaagcattcaaggggagcttgggtgctgtttaaagcattc 451
Db	490 GAAGAAATTTCTATCTCAGCAAGCAATTCAGAGGTGACTGGCTCTTTAAAGGCATTC 549
QY	452 agttcataaaggagagcagacataaagatcagaanaattgcacctgcacaatgtgata 511
Db	550 AGATTATATAGGGAAGCAGACGATATAAAGTTTCAGAAATTTGCACCTCGACATGTGATA 609
QY	512 aaaaaaagaaacccattctctgaggggaatcaagctggtcgtcagaaattgcatatgct 571
Db	610 GAGAGAAAGCCCATTTTGTGAGAGAAATTCAGGCCAGCTGCATTAATTTGCATTAAGT 669
QY	572 aatggagagcgtgaatgtttaatctctcaagaacaattgggaaataatatctctgtgacatgtcag 631
Db	670 AACCAAGAGCCCAATGTTAATCCCAAGACAAACGAGAGACATATCCAGGCAATATCAG 729
QY	632 aggtctcaacagcagctcatcaatcaatcactgtgcgtgagagcctcaggagaataatgttttgt 691
Db	730 AGGTTTCACACACACCCCTCCCGTCACAGGCCCAAGAGCCCTTAGAGAAATGTTTTCTT 789
QY	692 gggac-----aggcccaagggttcctctgtctgtgtgcagccttagagactgtgtgc 741

```
Db 790 GGACCTGTGCAGCCCTAGTACTTGTGACACCTGTGTGCCAGCTACTCCAGTAGTGCTGAA 849
Oy 742 ctgtgtccagttaa--ctcagctgtgtgtctcaagaggtgaaagcccaagccctgtgag 799
Db 850 AGGGGCAAAATATACACTTACGCTTACGCTTGTGAGGGGTGACAGGCGCAAGCCTTGGCAG 909
Oy 800 ctccaaagtgtgtgtgagcctgtgtgtgagaaagtcagaagaatgtgaggtctgtgagacct 859
Db 910 CTTCGGCATGTGTTGAGCAATGTCACAGAGCTCAGAAATTTGGATTGTGGAACT 969
Oy 860 ccaatcagaattcagaagatatatggaacccctgtgagtccagagcagaagttgtctgta 919
Db 970 CTGCCTAATTTTCAAGAGATGTACGGAAGTGTCCGATGCCAGCAAAAGTTTCTCTCA 1029
Oy 920 gagggtgtgtgtctcagagagacccctgtcaaggtgtgagtaaaagaggaagttgtgtg 979
Db 1030 GGGGCAAGGCGCTTATGAGAGAACTCTGTAGGGCAGTGTGGAAGGAAACATGGGGTCT 1089
Oy 980 gagcccccacacagagagtcaccagtggtgtccatctagtagagctgtgagaagagtc 1039
Db 1090 GAGCCCCACACAGAAATCTCTACTGAGGACACCTAGTGTGAGAGTGTGAGAAAGAGCA 1149
Oy 1040 ccatctctcagaagctccagaaggttagatccatcagctgacagctgtgacagatgtgtg 1099
Db 1150 CCAATCTCTCACACCCCAAGATATGATTCACCAACAGCTTGACCATGACACCTTGGA 1209
Oy 1100 tccacagaagctcagtgacagctgtgaaagcagcagagtgagtgctgtccctacaa 1159
Db 1210 GCCCGAGACATTCAACACACCCCTGAAAGCAGCCAAAGGGGGAATATATCTCGCAA 1269
Oy 1160 acagtagtgacagagctgacacagacgtgtggaatctacctctgtgacatgtgac 1219
Db 1270 GCCACAGGGGGAGGAGCTTCCCAAGACTATGGGAACCTACTCTTACATCAGACAGAC 1329
Oy 1220 gacgtgagacatgtgagtaaaagagatcatcttgagagctttaaagattgagctgcc 1279
Db 1330 TGTGTGACACTGTGAGTAAAGAGATCATTTTGGAGCTTAAATCTGACGTGCGCTCT 1389
Oy 1280 ggaattcgaactataltatgtggcc--cgtacccctgtgtgtgtgtgtgtgtgtgtgt 1338
Db 1390 GGATTTTCAAGACTTGCATGGGCGCTACACACCCCTTGTGTGGCGATTTTCCCAATTTGG 1449
Oy 1339 aactgcgatttaccacaaatgcgtgtaccctccattgtatgtag 1381
Db 1450 AAAGGCTGTATTATCCCAATACCTGTACCCCATTTGTATCTTG 1492
```

```
RESULT 3
LOCUS BE512633 1514 bp mRNA linear EST 11-OCT-2001
DEFINITION 32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION BE512633
VERSION BE512633.1 GI:16041645
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Lu, X., Cui, L. and Li, Y.
TITLE DDRT-PCR from B cell
JOURNAL Unpublished (2000)
COMMENT Contact: xingwu lu, liangxian Cui, yonghai Li
Department of Biochemistry
Institute of Basic Medical Science, Peking Union Medical College
Dongdan Santiao 5, Beijing, P.R.C, 100005
Tel: 86-010-65296951
Email: luxingwu263.net
full cDNA sequence.
```

FEATURES

Location/Qualifiers

Source

I..1514

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="human B lymphocyte cDNA library"
/tissue_type="bone marrow"
BASE COUNT 380 a 341 c 398 g 395 t
ORIGIN
Query Match 41.8%; Score 576.8; DB 10; Length 1514;
Best Local Similarity 81.0%; Pred. No. 1.2e-153;
Matches 737; Conservative 0; Mismatches 132; Indels 41; Gaps 4;
Oy 509 ataaaaaagaaacccattctcagaggaatcacaagctgtgtcagaaattgtcata 568
Db 1 ATAGAAAAGAAATCCCATATTTCTGAGGAGAAATCAAGCTGCGCAGAAATTTGCATA 60
Oy 569 tgaatgagagctgaaatgttaatctcaagaacatgggaaatattctctgtgacagt 628
Db 61 TGTAAAGAGATTAAATGTTATCCCAAGACAAATGGGAAATGTCTCCAGGGCATGC 120
Oy 629 cagaagcttcacagcagtcacatcaatcactgtgctgtgagagctagag--aaatg 685
Db 121 CAGAGCTTTCACAGACACCCCTCCATCAGAGCGCTGCAGAGCCCTAGAGAGAAAGTGA 180
Oy 686 ttgtgtggaagagccagaggtccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 745
Db 181 TCTCTTGGGCGAGGCGCCAGGGGTCCCATGCTGTGTGACGCTGGGAGCTTGTGCTTGT 240
Oy 746 gtcccaagttaattcaagc-----tgtagcttca 772
Db 241 GTCCAGCTGCTCCAGCATGGCTGTAAGGGGCCAACTAGAGCTCGGATTTGTGCTTCA 300
Oy 773 gagggtgcagcccccaagcctltgtgagcttccaaagtgtgtgtgtgtgtgtgtgtgt 832
Db 301 GAGGCTGCAAGCTGTAAACCTTGGCAGCTTCATGTGTTGAGCTGTGACCTGTGCAAG 360
Oy 833 aagtcagaatgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 892
Db 361 AAGTCAAAATTTGGGGTTTGGGAACTGTGCTTAATTTTCAGAAATGTGGAATGCC 420
Oy 893 tggatgccagcagaagttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 952
Db 421 TGGATGCCCTGGCAGAACTGTGCTGTAGGGAGGGGCGCTTCAATGGAATCTTCTTGG 480
Oy 953 gtagtacaagaaggaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1012
Db 481 GCAGTACAGAAAGGAAATGTGGGATCGGAGCCGCCAGCAGACAGAAATCCATCGGGC 540
Oy 1013 tctagtagactgtgtgagaagaagtcacacatccctcagactccagaaggtagatca 1072
Db 541 CCTAGTGGAGCTGTGAGAAAGGCGCACCAATCTCCAGACCCAGAAATGATGATCCACT 600
Oy 1073 gacagcttgacagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1132
Db 601 GACAGCTGTCTCATGACCTGTGAAACCCGAGA---CAACACACACCCCATGAAACCA 656
Oy 1133 gcaaggatgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1192
Db 657 GCTAGAGCAGAGAGGCTGTACCTGCAAAACCAAGGGGGAGAGCTGCCAAGATCATGGGA 716
Oy 1193 atctacctctgtcatltgtcatgtgacgtgtgagatgtgagatcaaaagatcat 1252
Db 717 ACCCACCTCTGATACATCAGATGACCTGTGATGAGATTGAGATTAAAGAGATCATTTT 776
Oy 1253 ggaagcttaagattgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1311
Db 777 GGAGCTTTAAAGATTGATGAGTCCCTCTAGATTGAGATTGAGTGGGCGCTGTAGCCCT 836
Oy 1312 ttgttttgccaatttttccatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1371
Db 837 TTGTTTACTCAATTTTCTCCATTTGGAGACAGCTGTGTTTACACAAATACCTGTACCCCA 896
Oy 1372 ttgtatgtag 1381
```

Db 897 TTGTATCTAG 906
|||||
RESULT 4 BM471183 1036 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT.6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049
DEFINITION 5', mRNA sequence.
ACCESSION BM471183
VERSION BM471183.1 GI:18520225
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1036)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM12293 row: g column: 18
High quality sequence stop: 661.
Location/Qualifiers
1..1036
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5563049"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPOrt6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 254 a 239 c 294 g 247 t 2 others
ORIGIN
Query Match 35.0%; Score 483; DB 10; Length 1036;
Best Local Similarity 77.7%; Pred. No. 7, 4e-127;
Matches 676; Conservative 0; Mismatches 151; Indels 43; Gaps 6;
QY 554 gcagaaattgcatatgtaatgagagct--gaatgtaatcctcaagaacaatgggaaa 611
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GCAGAAATTAATCAATAGTAGCAAGGAACTACTAATGTTATCCCAAGACATGGGAAA 60
QY 612 atatcccccggacatgtagagggcttcacagcgcttcacatcaatcagcgcttgaggc 671
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ATGTCTCCAGTTCTGTCAGAGACCTTCACGACCTCCATCCATCAGAGCCCAAGCG 120
QY 672 ctgagag--aaatggtttgttgagacagggctccctgctgtgtgacagccca 728
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CCAGGAGGAAAAAGTGTCTGTGTGGGCTTGAAGGTCCTCATCTGTGTGAAGCTTA 180
QY 729 gaagcttggtccctgtgtcccaatlaa----- 760
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GGGACTTGTCGCCCTGTGTCCAGCTGCTCCAGCCTTGCGTGAAGGGGCTAATGTATAG 240
QY 761 -----gcgtggtctcagaaggtgcaagcccaagccttgagcag---cttcaagtgtg 812
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CTCTGCTGTGCTTTAGAGGGTGTGAACCCCAAGCCTTGCGACTTCTCATGTGTG 300
QY 813 ttgagccttggtgccaagaagtaagaattgaggttgtagacctccaacagatttc 872
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 301 TTGACCTGTGGGTGCACAGAGTCAAGAACTGAGCTTTGGGACCTCTGCT-AGATTCC 359
QY 873 agaagatalatgaaacccctgtagtgccagcagaagtttgcttaggggtgggtcct 932
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 ACAAGATGATATGGAATAGCCCTGGATGGCCAGCAAAAGTTCTCTCAAGTTGGGGCCTT 419
QY 933 catggaagacctgtgcaaggtagtacaaaaggaaatgttggtgtggagcccccaca 992
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 CATGAGAAACCTCTGTAGGCACTGTGGAAGGAAATGTGGGCTTGAGCCCGCCACA 479
QY 993 ggtgtcccaagtggtgggtccatctagtagagcttgagaaagagtcacatccctcagaac 1052
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 GTGCCCTGTGGGCACTGCTAGTGAGCTGTGAAGAGAGGCCACCATCTCCAAAC 539
QY 1053 tccagaaggtatagatccactacagcttgagcatgtgctctgaaataccacaagacatc 1112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 540 CCCAGATATGTATGATGATCATCAACAGCTTGACATGCTGACTTGGAAGAGCCGAGACATC 599
QY 1113 agtgcagcctgtgaaagcagcaggaatgagtcgtaccctacaagacgtagtgagcag 1172
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 600 GACACCAGCCCATGAAAGCACACTGGAGGAGGCTATACCTGCAAGACACAGGGCGAG 659
QY 1173 agctgacaaagacggtgggaatctactctctgcatgtgcatgacgttgagagcagcag 1232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 660 AGCTGCTTAAGACCATGAGAACCCACCTCTTGTCATGACATGACCTGGATATGACACTTG 719
QY 1233 gagtcaaaagatcaatttgagacttaagattgactcccccactgattgagact 1292
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 720 GAGTCAAGAGAGATCATTTTGGACTTTGAATTTGATTTGCCCGCTGGATTTGACACTT 779
QY 1293 atatggggccgt-aacctgtgtttggccaattttccattggaagcgcgattt 1351
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 780 GTGTANGCCCTGATACCCCTTTGTTGGCAATTTCTCCATTTGAATGGCTACATTT 839
QY 1352 acccaatgctgttacctcactatgtagt 1381
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 840 TACCAATACCTGTACCCCTATGGATCTAG 869
RESULT 5
BI488505 870 bp mRNA linear EST 28-AUG-2001
LOCUS 603021014F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191803 5',
DEFINITION mRNA sequence.
ACCESSION BI488505
VERSION BI488505.1 GI:15327733
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 870)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1479 row: k column: 04
High quality sequence stop: 849.
Location/Qualifiers
1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5191803"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"


```
Db      857 CCATTGGGTGANA 869

RESULT  8
AG086951
LOCUS   AG086951 717 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence.
ACCESSION AG086951
VERSION   AG086951.1 GI:16638753
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
          BAC library clone:PTB-085H08.F.
ORGANISM Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Tokohi,Y., Watanabe,H. and Sakaki,Y.
TITLE    BAC end sequences of library PTB
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 717)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Tokohi,Y., Watanabe,H. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbes@sc.riken.go.jp, URL:http://nbp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
          Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the Rad process and may have higher chance of
          clone tracking errors.
COMMENT  PRIMERS
          Sequencing: -21M13
          LIBRARY
          Vector : pKS145
          R.Site 1 : SacI
          R.Site 2 : SacI.
FEATURES
          Location/Qualifiers
            source          1..717
                           /organism="Pan troglodytes"
                           /db_xref="taxon:9598"
                           /clone="PTB-085H08.F"
                           /sex="male"
                           /cell_type="lymphoblast"
                           /clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 175 a 189 c 195 g 158 t
ORIGIN

Query Match 31.2%; Score 430.4; DB 12; Length 717;
Best Local Similarity 82.4%; Pred. No. 7.2e-112;
Matches 506; Conservative 0; Mismatches 106; Indels 2; Gaps 1;

Qy 746 gtccagcaataatcagctgtggtcagaggggtgcaagcccaagcctgtgagcttcca 805
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 gtcccgagagctcagagctgtggtcagaggggtgcaagcccaagcctgtgagcttcca 149
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 806 agtggctgtgagcctgtggtgcaagaagtcagaattgaggttgggaacctccaatc 865
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 catgctgttaagccctgtagtgacacgaatctaaagactgggggttggaacctctgcct 209
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 866 agattcagaagatataggaacccctgtagtgcagcagaagaatttgcttagaggtg 925
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 agatttcagagagatattatgtatattacgttgatgccacagacaaattttgcttagggag 269
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 926 gggctcattaggaacctctgcaagggttagtacaagaagggaattgtgtggtggagcc 985
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 ggagcctcatgaaaaaactctgctgtagggcagtgtaggaagaaatgtgggctcagagcca 329
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 986 ccacacagaagtcaccagtggtggtctcatctagtagagctgtgtagaagaagtcaccatcc 1045
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db      330 TCACACAGAGTCCCTACTGCGGGCCACCCCTAGTGGAGCTGTGANAAGTGGCCACTATTTC 389
Qy 1046 tccagactccagaaggttagatccactgacagcttgcaagcatgtgctgtaaaatccaca 1105
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 390 tccagaccccaagatgtgtagtaccactgacagctgttcacacactgactggaaaaagccaca 449
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1106 gacacttagtccagcctctgaaagcagcagggatgtagtctgtacctacaacacgcta 1165
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 450 gacactcaatgcacagcctgtgaaagcagcagggatgtagtctgtacctacaacacacaca 509
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1166 gggagcagactgacagcagccgttgggaatcctctgttcattgtatattacccgagcgtg 1225
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 510 gggagcagactgaccccaagcagcagccacacacctgttcattatattacccgagcgtg 569
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1226 agacatgagtcacaagaagatcattttagcttgaattgagcttgcaccccaatgattc 1285
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 agacatgagtcacaagaagatcattttagcttgaattgagcttgcaccccaatgattc 629
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1286 cggacttataatggggccgta--ccctctgttggccaattttccatttgaactg 1343
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 630 cagacttgcatggggccctgtagccctttttttttggccaatttcccatgttggaatgg 689
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1344 ccgtatttaccaca 1357
    ||| | | | |
Db 690 ggtaattttaccaca 703
    ||| | | | |

RESULT  9
AG064424
LOCUS   AG064424 676 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-053H24.R, genomic survey sequence.
ACCESSION AG064424
VERSION   AG064424.1 GI:16616226
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
          BAC library clone:PTB-053H24.R.
ORGANISM Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Tokohi,Y., Watanabe,H. and Sakaki,Y.
TITLE    BAC end sequences of library PTB
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 676)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          Tokohi,Y., Watanabe,H. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbes@sc.riken.go.jp, URL:http://nbp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
          Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the Rad process and may have higher chance of
          clone tracking errors.
COMMENT  PRIMERS
          Sequencing: M13Rev
          LIBRARY
          Vector : pKS145
          R.Site 1 : SacI
          R.Site 2 : SacI.
FEATURES
          Location/Qualifiers
            source          1..676
                           /organism="Pan troglodytes"
                           /db_xref="taxon:9598"
                           /clone="PTB-053H24.R"
                           /sex="male"
                           /cell_type="lymphoblast"
                           /clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 164 a 165 c 191 g 156 t
ORIGIN
```


Query Match:	29.5%	Score 407.8	DB 10	Length 955
Best Local Similarity	76.9%	Pred. No. 2.3e-105		
Matches 645	Conservative	0	Mismatches 142	Indels 52
				Gaps 10
Qy 397	aattctcaacacaagaatcctcaagagtgatctggtggtgtgtttaaagcattcagttt	456		
Db 1	AATTCTCAACACACAGCATTAAGAGTGATCTGGTGTGTGTAAGCATTCAGTTT	60		
Qy 457	cataagggagcagagcatalaagagtttcagaataattgcacccctgcacatgtgataaaaa	516		
Db 61	TAAAGGAACACAGACATATAAGCTTGGAAGTTTGACGCTGACATGCAATGAGAAAA	120		
Qy 517	gaaaaaccccttcttcgaggggaaatcaagctgcgcgcagaatttgcatagtatga	576		
Db 121	GA AAA-----TTCAGAGTACATTTCAAGGACAGCTGCACAAGTTTGCAATGATATGA	172		
Qy 577	ggaagctgaatgttaacccccaagacaatctgggaaataatctccctgcagatctgaaggtc	636		
Db 173	GGGGCCAATGTTAATCCCAAGAAAAATGGGAAAAATGTTCTCCAGGGCATGTAGAGATC	232		
Qy 637	ttcacagagctccatcaataaactcactgccttggagggccttagag--aaatgtgtttg	694		
Db 233	TTCAATGCGACCCCTCCCATCAAGAGCCCTGCAGGGCTGAGGAAAAAGTGTTTCATG	292		
Qy 695	acaaagcccaaggtctccctgtgtgtgtgtcaagccttagagatctgtgtccctgttccagtt	754		
Db 293	GCGAGGCCCCAGGTCCCATGCTGTGTGAAGCCTTAGGGAGCTTGCGCCCTCATCTAGGCC	352		
Qy 755	aattcagc-----ttgtgctttaaagaggtga	781		
Db 353	GCTTAAGCCATGGCAAAAAGGGCCACAGCTAGAGCTCAGCATGTGGCTTTAGAGGGTGCA	412		
Qy 782	agccccaagccttgcagcttcccaagtgtgttgaacctgtgtgggttgcagaagaatcaga	841		
Db 413	AGCCCCAAGCCTTGGCAGCTT-CACATGSGTGTGAAGCCTGCAAGTGACATGTAGTCAGA	471		
Qy 842	atgaggtttgggaacctccaatcagattcagaagatatggaaaacccctgagatgcc	901		
Db 472	ACTGGGTTTGGGAACTCCACCTGATGATTTGAGAGATGTATGGAATGCTCTGGATGGCCC	531		
Qy 902	aggaagaagtttcttgaagggtggaggtcccatatgaagaaccttgaagaagtagt-aca	960		
Db 532	AGGCAGAA-TTTCTCTCATATGGGAGGGCCCTTATGAGAACTCTTACTAGGCAAGTGTA	590		
Qy 961	aaaaggaatgttgggttgggagccccacacagagttcccccagtggtgctcattcagtag	1020		
Db 591	AACGAGAAATGTAGGTGCGGTGCGCCCAACAGAGATTTCTTCTGAGACACATGGCTAATG	650		
Qy 1021	agctgtgagaagaagttccacatctcctcagactccagaaggtagatccatctgcagc--t	1079		
Db 651	AGCTGTAGAGAGAGGCGCAGCATCTCCAGATCCAGATGATGATGATCCATGACAGCTT	710		
Qy 1080	tgcagcagctggtccgtaaaatccaaagacatcagctgctgcagccgtgtaaagcagcaggga	1139		
Db 711	TGCACACAGGCTCGAAAAAGTCGACGACATTCAAGACAG-CGTGTAAGAGCGCA-GGA	768		
Qy 1140	ttgagctgtacacctcaaacacgtagtgcagagcttgcagaagacggtggaatctac	1198		
Db 769	GGGAATCTGACTGAGAGAC---AGGCTAGAACTGTGCCAACAACATGGGAACCAAC	824		
RESULT 13				
LOCUS	BG281182	760 bp	mrna	linear
DEFINITION	602402107f1 NIH_MGC_20 Homo sapiens cdna clone IMAGE:454448 5',			EST 21-FEB-2001
ACCESSION	BG281182			
VERSION	BG281182.1	GI:13030107		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 760)	NIH-MGC	http://mgc.ncl.nih.gov/.	
2	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	Contact: Robert Strausberg, Ph.D.	Email: cgarbs-r@mail.nih.gov
3	Tissue Procurement: ATCC/DCMP/DP	CDNA Library Preparation: Ling Hong/Rubin Laboratory	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	DNA Sequencing by: Incyte Genomics, Inc.
4	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov	Plate: LILCM1227	row: n	column: 01
5	High quality sequence stop: 760.	Location/Qualifiers	1..760	
6	/organism="Homo sapiens"	/db_xref="taxon:9606"	/clone="IMAGE:4544448"	/clone_lib="NIH-MGC-20"
7	/tissue_type="melanotic melanoma"	/lab_host="DH10B (phage-resistant)"	/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(5). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."	
8	BASE COUNT	211 a	181 c	219 g
9	ORIGIN	181 c	219 g	149 t
10	Query Match	29.4%	Score 405.4	DB 10; Length 760;
11	Best Local Similarity	77.5%	Pred. No. 1e-104;	
12	Matches 571; Conservative	0;	Mismatches 126;	Indels 40; Gaps 5;
13	444	aggcatcagttcattcaaaaggagagacataaagaftcagaaatttgcacctgaca	503	
14	2	AGGATTTCATTTTAAAGGA--AACAGCATAAAGTTCAGAAATTTCACCGCAG	59	
15	504	atgtgataaaaaaagaaccatttcttgaggggaaattcaagctgctgcagaattt	563	
16	60	ATGAGTAGAAAAAGAAAAACCATTTTGTGAGAAAAATTCAACCCAGTCAGAAATT	119	
17	564	gcatagttaatgagagctgaaatttaattccccaagaagaatggggaaatatctcgtg	623	
18	120	GCGAATATACCAAGAAAGCTTAATGTTAATCCCCAAGACATAGGGAATAATTCAGAGC	179	
19	624	catgtcagagttcttacacagcagtcacatcaaatcaactgcttgaggcctcctagag--aa	680	
20	180	CATGTCAAGACCTTACAGCAGCCCTCCATCATCAAGGCCACAGGCAAGAGGAAAA	239	
21	681	aatgttcttgaggacagcccaagggtccctgctgtgtgacgctcctagagctgtgc	740	
22	240	AGTGGTTTCATGCGGCGAGGCCAGGAGTCCATGCTGTGTATACCTAGGAGACTTGGTGC	299	
23	741	ccgtgtcccaagtaattca-----gctgtgg	767	
24	300	CTTGTAATCCAGCACTCCAGCATGCGTGAAGAGGCCAATGTAGAGCTCGGGGCTGTGG	359	
25	768	cttcagaaggtgcaagccccaagctcttgacagcttcaagaftggttgagccttgagggtc	827	
26	360	CTTCAGAGGGTGAACCCCAAGCTGTGGCAGCTTCCATGTGGTTCACGCTGCAGAGT	419	
27	828	caaaagaagtcagaattgagtttgggaacctccaatcagaattcagaagatatatgaa	887	
28	420	CACAGAGTCAAGTAATGAGGTTTGGGAACCTCCACCTAGATTTCAGAGGA--GTATGGAA	478	
29	888	acccttgatgcccaagcagaagttctgtctgaagggtggggctctcatatgagaacctgc	947	
30	479	ATGCTTGAAGCCCAAGAGTTTGGTGCAGGGCCAAAGGCTCTCAAGATTAACCTGTG	538	

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
U00001	1	EST	human.	Homo sapiens					
U00002	1	EST	human.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
U00003	1	EST	human.	Homo sapiens					
U00004	1	EST	human.	Homo sapiens					
U00005	1	EST	human.	Homo sapiens					
U00006	1	EST	human.	Homo sapiens					
U00007	1	EST	human.	Homo sapiens					
U00008	1	EST	human.	Homo sapiens					
U00009	1	EST	human.	Homo sapiens					
U00010	1	EST	human.	Homo sapiens					
U00011	1	EST	human.	Homo sapiens					
U00012	1	EST	human.	Homo sapiens					
U00013	1	EST	human.	Homo sapiens					
U00014	1	EST	human.	Homo sapiens					
U00015	1	EST	human.	Homo sapiens					
U00016	1	EST	human.	Homo sapiens					
U00017	1	EST	human.	Homo sapiens					
U00018	1	EST	human.	Homo sapiens					
U00019	1	EST	human.	Homo sapiens					
U00020	1	EST	human.	Homo sapiens					
U00021	1	EST	human.	Homo sapiens					
U00022	1	EST	human.	Homo sapiens					
U00023	1	EST	human.	Homo sapiens					
U00024	1	EST	human.	Homo sapiens					
U00025	1	EST	human.	Homo sapiens					
U00026	1	EST	human.	Homo sapiens					
U00027	1	EST	human.	Homo sapiens					
U00028	1	EST	human.	Homo sapiens					
U00029	1	EST	human.	Homo sapiens					
U00030	1	EST	human.	Homo sapiens					
U00031	1	EST	human.	Homo sapiens					
U00032	1	EST	human.	Homo sapiens					
U00033	1	EST	human.	Homo sapiens					
U00034	1	EST	human.	Homo sapiens					
U00035	1	EST	human.	Homo sapiens					
U00036	1	EST	human.	Homo sapiens					
U00037	1	EST	human.	Homo sapiens					
U00038	1	EST	human.	Homo sapiens					
U00039	1	EST	human.	Homo sapiens					
U00040	1	EST	human.	Homo sapiens					
U00041	1	EST	human.	Homo sapiens					
U00042	1	EST	human.	Homo sapiens					
U00043	1	EST	human.	Homo sapiens					
U00044	1	EST	human.	Homo sapiens					
U00045	1	EST	human.	Homo sapiens					
U00046	1	EST	human.	Homo sapiens					
U00047	1	EST	human.	Homo sapiens					
U00048	1	EST	human.	Homo sapiens					
U00049	1	EST	human.	Homo sapiens					
U00050	1	EST	human.	Homo sapiens					
U00051	1	EST	human.	Homo sapiens					
U00052	1	EST	human.	Homo sapiens					
U00053	1	EST	human.	Homo sapiens					
U00054	1	EST	human.						

FEATURES	SOURCE	Location/Qualifiers
		1. .736
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:5194028"
		/clone_1lb="NIH-MGC_114"
		/lab_host="DH10B"
		/note="Organ: brain; Vector: PCMV-SPOrt6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Innsbrgen). Research Genetics tracking code 019. Note: this is a NIH-MGC Library."
BASE COUNT		163 a
ORIGIN		196 c 185 g 192 t

	Query Match	Similarity	Best Local	Matches	Conservative	Score	DB	Length	Indels	Gaps
		29.3%	76.6%	0	133	404.6	10	736		
						1.8e-104			38	4
Qy	655	atcactgcctgagagcctagagag---aaatggtttgtgagacagcccaagttccct	711							
Db	733	ATCACAGACCCAGAGAGCTAGAGCCAAATAATGCTTTTGTGCTGTGACCCAGAGGTCCCC	674							

[illegible]

LOCUS	AG062447	RESULT 15
DEFINITION	AG062447	
ACCESSION	Pan troglodytes DNA, clone: PTB-050M20.F, genomic survey sequence.	
VERSION	AG062447	
KEYWORDS	AG062447.1 GI:16614249	
SOURCE	GSS: GSS (genome survey sequence).	
	Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male	
ORGANISM	BAC library clone: PTB-050M20.F.	
	Pan troglodytes	
	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.	
REFERENCE	1 (sites)	
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
	Totoki, Y., Watanabe, H. and Sakaki, Y.	
TITLE	BAC end sequences of library PTB	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 691)	
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,	
	Totoki, Y., Watanabe, H. and Sakaki, Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-AUG-2001) Aao Fujiyama, The Institute of Physical	
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC),	
	1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan	

COMMENT

(E-mail:chimpbes@sc.riken.go.jp, URL:http://nbp.gsc.riken.go.jp/,
Tel:81-43-503-9111, Fax:81-43-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES

source
1. .691
Location/Qualifiers
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-050M20.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 160 a 167 c 201 g 162 t 1 others
ORIGIN

Query Match

Best Local Similarity 82.1%; Score 404.2; DB 12; Length 691;

Mismatches 503; Conservative 0; Mismatches 103; Indels 7; Gaps 3;

```
QY 746 gtccagtaattcagctgtgtctcagaggtgtcaagcccaagccttgagccttcca 805
    ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 gtccggggagctcagagctgtgcttgcagagggtgcaagcccaagccttgagccttcca 142
QY 806 agtggatgtgaagcctgtgtgtgtgcaagaagtcaagaattgaggttgggaaccttccatc 865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 tttgggttgagacctgtgagtcacag----aagaatcggagtttgggaaccttccactt 197
QY 866 agattcagaagatatatgaaaccccttgatgcccagagcagaagtttgcgtgaagggtg 925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 agatttcagaagatgtatgaaatattacatgagtcgccatgcgaanaagtttgcgtgaggtca 257
QY 926 gggctctcatgtagaacctctgcgaaggtgtacaaagggaattgttgggtggagccc 985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 gggccctctatggagacttgcctagggcagctgtgaaaggaatgtgaggtcagagggcc 317
QY 986 ccacacagagtcaccagtgagggtccatctagtagagctgtgagaagaagttccaccatcc 1045
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 ccacacagagtcaccttactgagggcaccacacagcagctgtgagaagagggccactgtcc 377
QY 1046 tccagactccagaaggttagatccactgacagcttgcagcatgtgcctgaaaaatccaca 1105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 tccagactccagaatgttagatccactgagtgagctgtgcatcatgtgcttgaaaaagccaca 437
QY 1106 gaacctcagtgccagcctgtgaaagcagcagagtgatgagtcgttaccctcaaaaacgta 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438 gacactcaattgccacaccatgaaagcaccgggaagagagcttacccttacaagccacaca 497
QY 1166 gtgycagaagctg-accaagaacgltggaatctactcttgcattgcatgacctggaagt 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 498 ggggacagagctgttccagagaccatgggaacccaccttgcctcagtgagacctggtgt 557
QY 1225 gagacatgagtcacaagaagatcatcttggagctttaaagattgactgcccacatgatt 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 558 gagatattgagtcagagagatcatattgtagagcttttaagatttgactgtgtggctggatt 617
QY 1285 tggagcttatgagggcccgta-cccccttggtttggcacaatttccatttggaaactg 1343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 tgggaccttgcattggggccctgtagctcctttgttttggccaatttccatttggtaattg 677
QY 1344 ccgatttaccaca 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 678 ctgtatttaccaca 690
```

Search completed: June 30, 2002, 14:02:47
Job time: 8074 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2002, 12:22:13 ; Search time 1885.54 Seconds

(without alignments)
15326.934 Million cell updates/sec

Title: US-09-997-610-1

Perfect score: 1381

Sequence: 1 gatagtgtgcatcaccctgtctc.....tgtaacctcattgtatgtag 1381

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank : *
1: gb_da : *
2: gb_htg : *
3: gb_in : *
4: gb_cm : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pt : *
10: gb_ro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vi : *
15: em_da : *
16: em_fun : *
17: em_in : *
18: em_mu : *
19: em_mu : *
20: em_mu : *
21: em_or : *
22: em_ov : *
23: em_pat : *
24: em_ph : *
25: em_pl : *
26: em_ro : *
27: em_sts : *
28: em_un : *
29: em_vi : *
30: em_htg_hum : *
31: em_htg_inv : *
32: em_htg_other : *
33: em_htg_inv : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	1284	93.0	145880	9	HS302D9	Z82198 Human DNA s
2	697.6	50.5	198545	2	AC0117063	AC0117063 Homo sapi
3	693.4	50.2	124518	9	AL138962	AL138962 Homo sapi
4	693.2	50.2	77405	2	AL353634	AL353634 Homo sapi
5	692.6	50.2	40714	9	HSU212C1	Z69722 Human DNA s
6	690.8	50.0	123070	2	AC008799	AC008799 Homo sapi
7	690.8	49.9	189768	2	AC044889	AC044889 Homo sapi
8	689	49.9	154090	9	AC025577	AC025577 Homo sapi
9	689	49.9	186660	2	AC026107	AC026107 Homo sapi
10	686.6	49.7	170368	9	AC091982	AC091982 Homo sapi
11	686	49.7	67984	9	AC010312	AC010312 Homo sapi
12	684	49.5	168502	9	AC091005	AC091005 Homo sapi
13	682.4	49.4	173390	9	AC012038	AC012038 Homo sapi
14	681.6	49.4	131215	9	AC079614	AC079614 Homo sapi
15	681.4	49.3	173480	9	CNS00M8T	AL079343 Homo chr
16	681	49.3	38235	9	AC004559	AC004559 Homo sapi
17	681	49.3	146743	2	AC093588	AC093588 Homo sapi
18	681	49.3	166679	9	AC079899	AC079899 Homo sapi
19	681	49.3	176426	9	AC007370	AC007370 Homo sapi
20	680.6	49.3	152544	9	CNS05TEJ	AL359232 Homo chr
21	680.6	49.3	203726	2	AC011882	AC011882 Homo sapi
22	679.6	49.2	124531	9	HSJ633H17	AL049710 Homo sapi
23	679.2	49.2	187898	2	AC016715	AC016715 Homo sapi
24	679	49.2	138271	9	AC006360	AC006360 Homo sapi
25	679	49.2	147971	9	HS431P23	AL009178 Homo sapi
26	678.2	49.1	107885	9	AC006389	AC006389 Homo sapi
27	678.2	49.1	150332	9	AC004921	AC004921 Homo sapi
28	678	49.1	181842	2	AL391823	AL391823 Homo sapi
29	677.8	49.1	77774	9	AP000339	AP000339 Homo sapi
30	677.8	49.1	94730	9	AP000230	AP000230 Homo sapi
31	677.8	49.1	100000	9	AP000144	AP000144 Homo sapi
32	677.8	49.1	100000	9	AP000217	AP000217 Homo sapi
33	677.8	49.1	100634	9	AP001594	AP001594 Homo sapi
34	677.8	49.1	340000	9	AP001695	AP001695 Homo sapi
35	677.8	49.1	340000	9	AP001760	AP001760 Homo sapi
36	677.6	49.1	155764	9	AC011238	AC011238 Homo sapi
37	677.4	49.1	182972	2	AC023550	AC023550 Homo sapi
38	676.4	49.0	125295	2	AL672061	AL672061 Homo sapi
39	676.4	49.0	174662	9	AC026036	AC026036 Homo sapi
40	676.2	49.0	123631	9	HS22P01	AL109967 Homo sapi
41	676.2	49.0	168863	9	AC011286	AC011286 Homo sapi
42	676	49.0	105692	9	AL451046	AL451046 Homo sapi
43	676	49.0	135623	2	AC022618	AC022618 Homo sapi
44	675.8	48.9	23808	2	AC022460	AC022460 Homo sapi
45	675.4	48.9	150001	9	AC006063	AC006063 Homo sapi

ALIGNMENTS

RESULT	1	145880 bp	DNA	linear	PRI 12-DEC-1999
LOCUS	HS302D9	Human DNA sequence from clone RPI-302D9 on chromosome 22	Contains		
DEFINITION	Human DNA sequence from clone RPI-302D9 on chromosome 22	Contains			
ACCESSION	Z82198				
VERSION	Z82198.2	GI:6572207			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 145880)				
TITLE	Direct Submission				
JOURNAL	Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk				
COMMENT	On Dec 13, 1999 this sequence version replaced gi:3164067. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission				

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em, EMBL, SW, SWISSPROT, Tr, TREMBL, Wp, WORMPEP, Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/MGP/chr22>

RPL-302D9 is from the library RPCR-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RPL-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES

Location/Qualifiers

1..14580

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22"

/clone="RPL-302D9"

/clone.lib="RPCR-1"

188..245

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

FEATURES
source

Location/Qualifiers
1..40714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="LLOXNC01-212C1"
/clone_11p="LLOXNC01"
10..233
/note="L1 element fragment"
388..758
/note="L1 element fragment"
932..1108
/note="L1 element fragment"
2571..2617
/note="MER43 element fragment"
2613..5203
/note="L1 element fragment"
5204..5229
/note="13 copies of 2 mer 100 & conserved"
5230..5427
/note="MER43 element fragment"
8954..9088
/note="L1 element fragment"
12171..13985
/note="L1 element fragment"
14179..14612
/note="L1 element fragment"
14662..14780
/note="L1 element fragment"
14824..15044
/note="L1 element fragment"
15234..15369
/note="L1 element fragment"
15421..15849
/note="L1 element fragment"
16026..16155
/note="MSTR element fragment"
16073..16149
/note="THEIB element fragment"
16204..16495
/note="Alu repeat: matches 1..308 of consensus"
16532..16609
/note="THEIB element fragment"
16624..16722
/note="MSTR element fragment"
16626..16737
/note="MSTR element fragment"
16651..16712
/note="THEIB element fragment"
17060..17653
/note="L1 element fragment"
19720..19797
/note="MER21B element fragment"
19877..19915
/note="MER21B element fragment"
20007..20256
/note="MER21B element fragment"
20518..20734
/note="MER30 element fragment"
22651..22730
/note="MLT1A element fragment"
22740..22814
/note="MLT1A element fragment"
22917..22974
/note="MLT1A element fragment"
23342..23501
/note="THEIB element fragment"
23550..23697
/note="THEIB element fragment"
23725..24279
/note="THR element fragment"
24395..24593

repeat_region
24643..25270
/note="THR element fragment"
repeat_region
25277..25463
/note="THR element fragment"
repeat_region
25528..25627
/note="THEIB element fragment"
repeat_region
26081..26196
/note="THEIB element fragment"
repeat_region
26265..26661
/note="MLT1A element fragment"
repeat_region
26797..26898
/note="LTR2 element fragment"
repeat_region
27949..28241
/note="MSTR element fragment"
repeat_region
28319..28608
/note="Alu repeat: matches 308..1 of consensus"
repeat_region
30032..30159
/note="Alu repeat: matches 308..1 of consensus"
repeat_region
31794..31980
/note="MER5 element fragment"
repeat_region
31981..32073
/note="Alu repeat: matches 308..109 of consensus"
repeat_region
32232..32735
/note="Alu repeat: matches 93..1 of consensus"
repeat_region
35168..35654
/note="MER9 element fragment"
repeat_region
40640..40693
/note="MER4B element fragment"
repeat_region
40640..40693
/note="27 copies of 2 mer 94 & conserved"
BASE COUNT 11544 a 7879 c 8141 g 13150 t
ORIGIN

Query Match 50.2%; Score 692.6; DB 9; Length 40714;
Best Local Similarity 80.6%; Pred. No. 1.8e-191;
Matches 873; Conservative 0; Mismatches 174; Indels 36; Gaps 4;

QY 335 ctgcatactatgtcaagtgatataatttggcctaaggaagcaaatlttgctataagaa 394
|||||
DB 24003 CTGCCCTAGAGATCTGTGGAAATTTGACCTTGAGAGAGATTTGAGTATGAGTGTA 24062
QY 395 gaatttccaagcaagaagcatccaagagtgactgggtgctgttaagagcatcaat 454
|||||
DB 24063 GAAATTTCTAACGACGAAGCATCTCAAGAGTGACTGTGTCTGTAAAGCAGCTCACT 24122
QY 455 ttcataagggagcagagataagagttcagaanaatttgacccttgacaaatgtataaaa 514
|||||
DB 24123 TTTCGAAAGGAGAGCAAGAACATTAAGATTTCAGAAATTTTCACCTGCACAAATGTGATAGAA 24182
QY 515 aagaaanaaccca-ttttcgagggaggaatccaagctgggtgctgaagaatttgcatatgtaa 573
|||||
DB 24183 AAGAAACCCCATTTTTCGATGATTAATGAAGCCAGCTGGAGAAATTTGATTAAGTAA 24242
QY 574 tgaagagctgaatgttaattccatccaagacaatggggaanaatctccctgacatgcaag 633
|||||
DB 24243 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGAAATATCTCATGTGCTGACAG 24302
QY 634 gtcttcacagcagtcacatcaaatcaactg-gcctggagggcctaggaanaatgtttgtg 692
|||||
DB 24303 GCTTCACGGGAGCCCTCATCACTACATGCCCCAGAGGCTGAGAAATGGTTTATG 24362
QY 693 ggaagagcccaaggtccctgtgctgtgcaagcctagagaactgtgtccctgtgccag 752
|||||
DB 24363 GGTGGGGCCCAAGGTCCTCATCTATGTGCTAAGGAGATTTGGTCCCTGCTCCAG 24422
QY 753 ttaattca-----gcttgagcttaagaggtg 779
|||||
DB 24423 CCATCTCACGACCATTAACGTGAAGGGCCAAAGATATAGCTTGGGCTGTGCTTCAAGGATG 24482

```
Qy 780 caagcccaagccttgagcagcttccaagtggctgtgtgagcctgtggtgcaagaagctca 839
    |||||
Db 24483 GAAGCCCCAGGGCTTGAGAGCTTCATCGTGTGGAGCTGCAGAGTGCATGAACTCAA 24542
Qy 840 gaattgaagtttggaacctccatcagatttcagaagatataatgaaacccctgattgc 899
    |||||
Db 24543 GAATTAGGTTTGGGAACCTCCACCTTAATTTCAAGAAATGTAAGAAATGCTCGATGC 24602
Qy 900 ccagcagaagtttgctgtaggggtggtggtctcattatgaaagaccttcgaagtgatgac 959
    |||||
Db 24603 CCAGCGAAAGTTTGCTGAGGGGGGAGAGACCTCATGAGACACTCGTTAGGGTGTGT 24662
Qy 960 aaaaaggaaatgttggtgagagagccccaacagagatccccaagtgagggtcactagta 1019
    |||||
Db 24663 GGAAGGGAAGATGGGGTCATAGCGCCACACAGAGTCCTTACCAGGGGACCACTAGTG 24722
Qy 1020 gaactgtgagaagaagtcacatctccacatccacagaaaggtatagatccactgagact 1079
    |||||
Db 24723 GAGCTTGAGAAAGAGGCCACCATCTCCAGACCCAGAAATGGTAGATCCACGACAGCC 24782
Qy 1080 tgcagatgtgctgtaaaatacaacagacactagtcgagcctgtgaaagcagacagga 1139
    |||||
Db 24783 TGACACGCTGCTGGTGAAGAGCCGAGACACTCAATGCAAGCCTGTGAAAGCAGAGGA 24842
Qy 1140 tggagtcgtaccctacaacacgtagtgtgagagctgcgaccagaacgtgaggatctacc 1199
    |||||
Db 24843 GGGAGGCTGTACCTCGCAAGACACAGGGGCGAGCTGCCAAGATCATGGGAACCAATC 24902
Qy 1200 tcttgatgtcatgacactgagcagctgagaacatgagtcacaaagagatcatttgagact 1259
    |||||
Db 24903 TCTTGATACGATGACTCGATGATGAGACATGAGATGCAAGAAATCATTTTGGAGCTT 24962
Qy 1260 taagatttgcctgcccacactgatttcggaacttataatgagggcccgta-cccttggttt 1318
    |||||
Db 24963 TAAGATTGACGTGCTCCACTGGATCTCAGACTTCATGAGGGGCGCTGAGCCCTTTGT 25022
Qy 1319 ggcgaattttcccatcttggaactgcgatttaccccaatgctgtcacttcactgtatg 1378
    |||||
Db 25023 GGGCAATATCTCCATTTGGAATGGCTGTATTATTCCAATGCCCTGTAACCCCATTTGTATC 25082
Qy 1379 taq 1381
    |||
Db 25083 TAG 25085

RESULT 6
AC008799 123070 bp DNA linear HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 5 clone CTP-2061E19, WORKING DRAFT
DEFINITION
SEQUENCE, 8 ordered pieces.
ACCESSION AC008799 GI:9256046
VERSION AC008799.4
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 123070)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 123070)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
COMMENT On Jul 18, 2000 this sequence version replaced gi:7709316.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
```

```
Center Project Name: 651088
Center clone name: CTRB-HL_2061E19
-----
Summary Statistics
Consensus quality: 116295 bases at least Q40
Consensus quality: 121288 bases at least Q30
Consensus quality: 122086 bases at least Q20
Estimated insert size: 123000; pulse field gel estimation
Estimated insert size: 122770; sum-of-contrigs estimation
Quality coverage: 6.04 in Q20 bases; pulse field gel estimation
Quality coverage: 6.05 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contrigs. Gaps between the contrigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1
* 9437 9436: contig of 9436 bp in length
* 9537 26537: contig of 17001 bp in length
* 26538 26537: gap of unknown length
* 26538 26537: gap of unknown length
* 45958 45958: contig of 19321 bp in length
* 46058 46058: gap of unknown length
* 46059 66857: gap of 22599 bp in length
* 68658 68657: gap of unknown length
* 68758 72152: contig of 3395 bp in length
* 72153 72152: gap of unknown length
* 72253 81069: contig of 8817 bp in length
* 81070 81169: gap of unknown length
* 81170 121547: contig of 40378 bp in length
* 121548 121647: gap of unknown length
* 121648 123070: contig of 1423 bp in length.
FEATURES
    source          1..123070
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTP-2061E19"
                    /clone_lib="Caltech human BAC library D"
BASE COUNT  37763 a 24223 c 23855 g 36528 t 701 others
ORIGIN
Query Match 50.0%: Score 690.8; DB 2: Length 123070;
Best Local Similarity 80.0%: Pred. No. 6.5e-191;
Matches 853; Conservative 0; Mismatches 177; Indels 36; Gaps 2;
Qy 352 gaataatttgctaaatgaggaagcaaatlttgctaaataaggaagaaatttcagacga 411
    |||||
Db 107906 GAACCTTTAACTTGAAGAAATGATTAGGGTATCTGGCAGCAAGAAATTTCTAAGCACA 107965
Qy 412 aagcaattcaagaggtgacttggtgctgttaagggcattcaatcattataaggaagcaga 471
    |||||
Db 107966 AAGCATTCAGAGGGTGACTTGGGTCTTAAAGGTAATTCAGTTTAAAGGGAAGCA 108025
Qy 472 gataaagatcagaanaatttgcacccctgacaatgtgatataaagaanaaaccaatttc 531
    |||||
Db 108026 GCATTAAGTCTGGAATTTGCACTGCCACATGTGATGAAAGAAATTCATTTTC 108085
Qy 532 tgaagggaattcaagctgctgcagaanaatttgcataatgaaatgagagctgaaatgtaa 591
    |||||
Db 108086 TGAGGAGAAATCAAGCTGCTGCAGAAATGTGCATTAAGTAACAAAGAGACTGAATATTA 108145
Qy 592 tctcaagaacatggtgggaaataatctccctgacatgtcagaggtcttcacagactcat 651
    |||||
Db 108146 TCCCAAGACATGGGGGAAACATCTCAAGGCATGTCAAGAGTCTTCAAGCCACCCCT 108205
Qy 652 caatcagcgtcagagcctagagagaanaa---tggttttgaggagcagcagggctc 708
    |||||
Db 108206 CCATCATTAACCCCAAGAGCTTAGGAGGAAAGAGTGTGTTGTGGGCGAGGCCAGGGTC 108265
```


*	15528	15627:	gap of	100	bp
*	18628	20561:	contig of 1934	bp	in length
*	20562	20661:	gap of	100	bp
*	20662	22484:	contig of 1822	bp	in length
*	22485	22584:	gap of	100	bp
*	22585	22587:	contig of 2703	bp	in length
*	22588	22587:	gap of	100	bp
*	22588	227603:	contig of 2216	bp	in length
*	27604	27703:	gap of	100	bp
*	27704	30020:	contig of 2317	bp	in length
*	30021	30120:	gap of	100	bp
*	30121	33472:	contig of 3352	bp	in length
*	33473	33572:	gap of	100	bp
*	33573	38084:	contig of 4512	bp	in length
*	38085	38184:	gap of	100	bp
*	38185	41069:	contig of 2883	bp	in length
*	41070	41169:	gap of	100	bp
*	41170	45266:	contig of 4097	bp	in length
*	45267	45366:	gap of	100	bp
*	45367	51092:	contig of 5726	bp	in length
*	51093	51192:	gap of	100	bp
*	51193	55262:	contig of 4070	bp	in length
*	55263	55362:	gap of	100	bp
*	55363	60982:	contig of 5620	bp	in length
*	60983	61082:	gap of	100	bp
*	61083	67628:	contig of 6546	bp	in length
*	67629	67728:	gap of	100	bp
*	67729	72690:	contig of 4962	bp	in length
*	72691	72790:	gap of	100	bp
*	72791	79286:	contig of 6496	bp	in length
*	79287	79386:	gap of	100	bp
*	79387	83451:	contig of 4065	bp	in length
*	83452	83551:	gap of	100	bp
*	83552	92170:	contig of 8619	bp	in length
*	92171	92270:	gap of	100	bp
*	92271	96333:	contig of 4063	bp	in length
*	96334	96433:	gap of	100	bp
*	96434	103218:	contig of 6786	bp	in length
*	103219	103318:	gap of	100	bp
*	103319	112553:	contig of 9235	bp	in length
*	112554	112653:	gap of	100	bp
*	112654	123339:	contig of 10586	bp	in length
*	123340	123340:	gap of	100	bp
*	123340	137921:	contig of 14582	bp	in length
*	137922	138021:	gap of	100	bp
*	138022	157913:	contig of 19982	bp	in length
*	157914	158013:	gap of	100	bp
*	158014	189768:	contig of 31755	bp	in length

FEATURES	Location/Qualifiers
source	1..189768
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-792C1"
misc_feature	1..1015
	/clone.lib="RPC1-11 Human Male BAC"
	/note="assembly_fragment"
misc_feature	1116..1284
	/note="assembly_fragment"
misc_feature	2385..3898
	/note="assembly_fragment"
misc_feature	3999..5259
	/note="assembly_fragment"
misc_feature	5360..6912
	/note="assembly_fragment"
misc_feature	7013..8145
	/note="assembly_fragment"
misc_feature	8246..9694
	/note="assembly_fragment"
misc_feature	9795..11004
	/note="assembly_fragment"
misc_feature	11105..12907

```
misc_feature /note="assembly_fragment" 13008. .14916
misc_feature /note="assembly_fragment" 15017. .17114
misc_feature /note="assembly_fragment" 17215. .18527
misc_feature /note="assembly_fragment" 18628. .20561
misc_feature /note="assembly_fragment" 20662. .22484
misc_feature /note="assembly_fragment" 22585. .25287
misc_feature /note="assembly_fragment" 25388. .27603
misc_feature /note="assembly_fragment" 27704. .30020
misc_feature /note="assembly_fragment" 30121. .33472
misc_feature /note="assembly_fragment" 33573. .38084
misc_feature /note="assembly_fragment" 38185. .41069
misc_feature /note="assembly_fragment" 41170. .45266
misc_feature /note="assembly_fragment" 45367. .51092
misc_feature /note="assembly_fragment clone_end:T7
vector_side:right"
misc_feature 51193. .55262
misc_feature /note="assembly_fragment" 55363. .60982
misc_feature /note="assembly_fragment"
```

Query Match	50.0%;	Score 690.8;	DB 2;	Length 189768;
Best Local Similarity	80.0%;	Pred. No. 6,7/e-191;		
Matches 853;	Conservative	0;	Mismatches 177;	Indels 36; Gaps 2;
Oy 352 gaacattggcctatgaggagaagcaaatlttggcctatagaagaagaatttciaagcaga				411
Db 86586 GAACCTTTAACTTGAAAGAAATGATTTAGGGATCTGGCAGACAAATTTCTTAGCAGCA				86537
Oy 412 aagcattcaagaagtgactgtgctgtgttaaagccatcagtticaaaaggaaggcaga				471
Db 86536 AAGCATTCACAGAGTGACTGGGTGCTGTAAAGTATTCAGTTTATTAAGGAACAGA				86477
Oy 472 gcataaagaagtcagaanaatttgcacccctgcagcaatgtgtataaagaagaacccatttc				531
Db 86476 GCATPAAAGTCTGGAAATTTGCAATCCGACAAATGTATTAAGAAAGAAATTTCAATTTTC				86417
Oy 532 tgaagggaattcaagctgcgtgcagaanaattgtcatagtatagaagagctgaattca				591
Db 86416 TGAGAGAAATTCACACTGGCTGCAGAAATGTGCATTAAGTACACAGAGCTGAATTTTA				86357
Oy 552 tccctcaagacaatlgggggaanaatatcctcctgcagacatgtcagaaggtcttcacagcagtcac				651
Db 86356 TCCCCAGACAAATGGGGAAACATCTCCAGGACATGTACAGAGGTCTTCACGGAGCCCT				86297
Oy 652 caatcacctgcctcgtgagagcctagaagaana---tgatttctgttgagacagagcccaagctc				708
Db 86296 CCCATCTTAAACCCCAAGGCCCTAGGAGGAAAAAGTGTGTGTGGCCAGGCCCAAGGCTC				86237
Oy 709 cctgtgtgtgtgcacagccctagagactttgttgccctgtgtccacagttaatca-----				760
Db 86236 CTTTGTCTGTGTGCACCTTAGGGACTGTGGTGCCTGCATCCCACTGTCTCCACCATGGC				86177
Oy 761 -----gcctgtgtctcagaaggtgcagaagcccaagcctct				795
Db 86176 TGAAGGGGCCCACTAAGAGCTCGGGCTGTGGCTTCCAAAGGTCAAGCCCAAGCCTTG				86117
Oy 796 gcaagcttcacagtgtgtctgaagcctgtgtgtgcataaagaatlcagaanaattgaagtttggga				855
Db 86116 GCACCTTCATGTGGTGTGAGCTCTGCAGAGTGCACAGAAAGTCAAAATTTGGGTTTGGGA				86057

Oy	8556	acctcaatcgaagattcgaagaatataatgaatacccttgatcgcgaagcaagaatttc	915
Db	86056	acctccactcagatttgacagatgatcttatagaattgcccgatgcaccagcacaagcttcc	85997
Oy	916	tgtaggggttgagggttcctcaatgagaacctctgcaagggtagtlacaaaagggaattgttg	975
Db	85996	tgcacggggcagaggccctttatgacgaacctctctaggggcagtcgcaaaataattgtacg	85937
Oy	976	gtggagagcccccacacagagttcccgatgggtgcacatctagttagagctcgttagaagaag	1035
Db	85936	gtcacagggccacacacagagtaaccacactagagacactgccttattgagagctgtgagaagaag	85877
Oy	1036	tccacacatccctcagacctcagaaggtagatccacatgacatctgcagcatgtgcctcta	1095
Db	85876	gccacacatctctccacacatcccaaaaggtagagtcacacacacactctgacatgtgctactcttg	85817
Oy	1096	aaatcccaacagacatcagttgcagcgcctgtgaagcagcagaggtatggatctgtacccta	1155
Db	85816	aaaacccacagagcactctatttgccacgctctgtaaaagcagcagagagagcctatcccttg	85757
Oy	1156	caaaacgtagtgcagagctctgacccaagacgcttggaatctcaacctcttgcatgtcatga	1215
Db	85756	caaatctcacagggcggaagctctccaaagccttgggaaacctctcttgcctcacagcgtga	85697
Oy	1216	cctggaacgtgagaacatgagtgcaaaaagatcatcttggagctttaagaattgacctgccc	1275
Db	85696	cccggaattgaaacatgagtgctaaagagaaatattttggagcttttgaagatttgacctggcc	85637
Oy	1276	cactggaattcaggaactataatggggcccgctaacaccttggttcttgcgcaatttctccat	1335
Db	85636	tgcctggaatttcagactgcataggggccctgagcccttgcctttggcgcaattttctccattt	85577
Oy	1336	tggaaatgcgatttaccacaatgcctcttacctcatgtatgtag 1381	
Db	85576	tggaaacactgtattttaccctccatgcctctgtaaccctcagattgtactacttgg 85531	

RESULT		8
AC025577		
LOCUS		
DEFINITION	AC025577 Homo sapiens 12 BAC RP11-13C3 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	154090 bp DNA linear PRI 25-AUG-2000
ACCSSION	AC025577	
VERSION	AC025577.15	
KEYWORDS	HG. GI:9910028	
SOURCE	human.	
ORGANISM	Homo sap lens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiade; Hominio. Homo. 1 (bases 1 to 154090)	
AUTHORS	Muzny,D.M., Adams,C., Adio-Odofole,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Ameratunga,H.C., Are,J.R., Banks,T., Barbara,I.J., Benton,J., Blamege,K., Blankenburg,K., Bonnard,L., Bouk,Y., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunhay,C., Burckel,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,R., Cavazos,S.R., Catcher,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,D., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frintz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guenara,W., Gunartne,P., Hale,S., Hamillon,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Collins,B., Hornst,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliviet,S., Joudsh,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C., Kraticic,J., Kuresht,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Litcharge,O., Lieu,C., Liu,J., Liu,W.,	

TITLE JOURNAL
REFERENCE 2 (bases 1 to 1540930)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 1540930)
REFERENCE Worley,K.C.
AUTHORS Direct Submission
JOURNAL Submitted (25-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Aug 25, 2000 this sequence version replaced gi:9664948.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STIS are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats were identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 154090
Phrap values in estimate: 153545
Average error rate (BCM-Phrap estimate): 0.000261755
Fraction of Phrap values less than 40 : 0.02840086
Number of consensus changing edits: 22
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
9347 aacacagacc(n)ttttttttt aacacagacc(t)ttttttttt
10135 attgcccct(n)taagaaaga attgcccct(t)taagaaaga
10135 ctgcacacag(n)aaagatcca ctgcacacag(t)aaagatcca
10417 agcaatgcag(n)ggctacagaa agcaatgcag(t)ggctacagaa
61959 cttactatt(n)ttttattct cttactatt(t)ttttattct
61960 ttactattn(t)gtttattcta ttactattn(g)ttttattcta
61961 ttactattn(g)ttttattcta ttactattn(t)ttttattcta
61982 acaacacaag(n)ntaggttttg acaacacaag(a)ntaggttttg
61983 caacacaagn(n)taggtttg caacacaagn(t)taggtttg
61993 nttaggtttg(n)ttaactttcc nttaggtttg(t)ttaactttcc
62205 tcaatgacc(n)ctgttcacac tcaatgacc(c)ctgttcacac
62956 actgcaacct(n)tgctcccgac actgcaacct(c)tgctcccgac
63722 ttacatata(n)cagtaacagta ttacatata(e)cagtaacagta
90212 cctagaanaa(n)gaacttttct cctagaanaa(t)gaacttttct
90216 gaaaanaagac(n)tttcttttt gaaaanaagac(t)tttcttttt
90226 ttcttttta(n)tttttcat ttcttttta(t)tttttcat
90236 taatttttc(n)atcttaagg taatttttc(t)atcttaagg
90443 acccttagatg(n)cttcctccag acccttagatg(a)cttcctccag
91405 cacagcttaa(n)caaggtaaag cacagcttaa(a)caaggtaaag
122290 aaaaagaag(n)gtaaggaaga aaaaagaag(a)gtaaggaaga
137090 acagagaag(n)caaaacacac acagagaag(a)caaaacacac
141632 tgaagcaact(n)gaaaagtaat tgaagcaact(t)gaaaagtaat

----- Distribution of Quality < 40 Bases -----
bases
10001 * * * * *
9001 * * * * *
8001 * * * * *
7001 * * * * *
6001 * * * * *
5001 * * * * *
4001 * * * * *
3001 * * * * *
2001 * * * * *
1001 * * * * *
01 * * * * *
5 10 15 20 25 30 35 40
Phrap Value Range

Version: 1.01 gxf.
Location/Qualifiers
source 1..154090
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-13C3"
54..288
repeat_region /rpl_family="MIR"
1020..1146
repeat_region /rpl_family="MER3"
complement(1248..1584)
repeat_region /rpl_family="MER46B"
complement(1624..1682)
repeat_region /rpl_family="MIR"
2630..2761
repeat_region /rpl_family="GA-rich"
4088..4216

/rpl_family="L2"
repeat_region 5232..5301
/rpl_family="MER102"
repeat_region complement(5450..5749)
/rpl_family="AluY"
repeat_region 6571..6616
/rpl_family="(Tg)n"
repeat_region complement(6678..7320)
/rpl_family="L1ME3"
repeat_region 7419..7454
/rpl_family="AT-rich"
repeat_region complement(7741..7807)
/rpl_family="MER5A"
repeat_region 8113..8267
/rpl_family="MER5A"
repeat_region 8385..8587
/rpl_family="MIR"
repeat_region 8589..8630
/rpl_family="L2"
repeat_region 8666..8695
/rpl_family="(TTA)n"

Query Match 49.9% Score 689; DB 9; Length 154090;
Best Local Similarity 80.1%; Pred. No. 2.2e-190;
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

Qy 352 gataatttgctaaaggaagcaaatlttgctaaataaggaagaaatttctaagcaga 411
Db 101594 GAACCTTGAACTTGAGACATGATTTAGGGTATCTGGCGGAAGAAATTTTAAGTACA 101653
Qy 412 aagcattcaagagtgagctgtggtcgttaagagcattcaagttcataagggagagcaga 471
Db 101654 AAGCATTAAGAAAGGAGTGTGATGATGATTAAGGATTCAGTTATATATAGGAGAGCAGA 101713
Qy 472 gcaataagagcttcagaanaatttgcacccctgcacaatgtgcataaanaaagaaacacatttc 531
Db 101714 GCATTAACCTTTGGAAATTTGACGCTGACATGTGATGAAAGAAAGAAACCCATTTC 101773
Qy 532 taaggggaaattcaagctgcgtcagaaatttgcataatgtaabgagagctgaattgttaa 591
Db 101774 TAGGGGAAATTTGAAGCTGCGTGCAGAAATTTGCATTAAGTAACAAGAGACCTTAATGTAAA 101833
Qy 592 tectcaagaacatgggggaaatactctcttgacatgtctcagagcttcacagatccat 651
Db 101834 TCCCCAGACAAATGGGGAATGTCCTCAAGGCATATCAAGGCTTCTCAATGGCACGCCAT 101893
Qy 652 caatcaactgagccttgagagccttagagaaatgttttgytgcagagcccgaggtccct 711
Db 101894 CCCATCACAGGCCCGGAGGCGCTTAGAGAGAAATGCTTTTGTAGGCCAGGCCAGAGCCCC 101953
Qy 712 gtgcgtgtgcagccttagagactgtgtccctgtgt----- 747
Db 101954 ATGCTGTGTGACGCTTAGGAGCTGTGTGCTGCTGTGTATACACAGCTCTGCTAGTGCTGCT 102013
Qy 748 -----cccaagttaattcaagctgtgtgtctcagaaggtgtcagaagcccaagccttgg 796
Db 102014 GAAAGGGGCAACATGACGCTTGCGGCGCTGCGAGGCTGTCAGGCCCAAGCCTTGG 102073
Qy 797 cagcttcacaagtggtgtgagcctgtgtgcagaagaatgcaagaattgaagtttgggaa 856
Db 102074 CAATTTCCATGTGTGTGACCTTAGAGGTGACACAGATGCAAGTAATGAGGCTTGGGAG 102133
Qy 857 cctcaatcagaattcagaagaatataatgaaaccccttgatgtccagagcagaagtttgt 916
Db 102134 CCGCACACCTAGATTTCAAGAGATGTATGAAATGCGCGATGTCACGAGAAATTTGCT 102193
Qy 917 gtaagggtgaggtctctcaatgagaacctgcagaaggtgagtlacaaaagggaagtgtggg 976
Db 102194 TAGAGGCTGGGCGCTCATGAGAACTGTGATGAGGCACTGCAGAGGAAGGAATGTGGGG 102253
Qy 977 tggggagcccccacagaaggtcccaagtgagggtcccaatctagtagagctgtgtgagaagt 1036
Db 102254 TTGAGACCCATACACAGATCCCTACTGTGGGGCACACATTAAGGGGAGCTGTGAGAGAGGG 102313

QY	1037	ccacatctctccagactccagaagggatgacatctggcagcttcgaagctggcgttaa	1096
Db	102314	CCACCATCTCTCAGACCCCAAGATGTGATTCACCAACAGCTTCCACCATGTACTTGGAA	102373
QY	1097	aatccacacagacactcagctggccagccctg tgaagcagcagagatlgagctctaccctaac	1156
Db	102374	AAAGCGACAGAGACTCAACACCAAGCCCATGAAAGACGCCAAGAGGAGACCTGTACCCTGCG	102433
QY	1157	aaaaccgtagtctggcagagctgcaccaaagccgttgggaactaactcttcgatctgcatgac	1216
Db	102434	AAAGCCACAGAAGACGAGACTTCCCAAGACATATGGGAACCAACCTCTGTGATCAGTGTGAC	102493
QY	1217	ctgagcgtlgagacatctgagtcataaagaagatcaatttggagcttaagaatttgaactgcgcc	1276
Db	102494	CTAGATGTGAAGAACATGCGTCAAAAGAGATCATATTAGAACATTTGAGATTGTGACTAACCT	102553
QY	1277	actggattctggactataatgagggcccgta -cccccttgcttttggccaatttttccatt	1335
Db	102554	TCTGGGTTTCAGACTTGCATGGAGGGGCAATPATAGCCCCCTTTGTGTGGCAATTTCTCCCAAT	102613
QY	1336	tggaactgcgctatttacccaatgcgcttacctcatgta	1376
Db	102614	TAGAACAAGCTGTATTACTCAATGCTGTATATGCGCATTTGCA	102654

RESULT	9
LOCUS	AC026107
DEFINITION	Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT
ACCESSION	AC026107
VERSION	AC026107.22 GI:13899178
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFLOT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 186660) Muzny,D.M., Adams,C., Adio-Oduola,B., All-oman,F.R., Allen,C., Alshrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Birmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gebisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrill,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,N., Holloway,C., Hollins,B., Homs,J.F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivel,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,U., Kovar,C., Kratochvic,J., Kunesht,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louiseded,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Meshawari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Messer,E., Mawhney,E., Mcloed,M.P., Meador,M., Mei,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Monabbot,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwaku,S., Ogut,M., Okunodu,G., Otungye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pichens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodajren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swatek,A., Tabor,T., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tusamti,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,U., Zorrilla,S., Nelson.D.,
Weinstock,G. and Gibbs.R.
Unpublished
Direct Submission
2 (bases 1 to 186660)

AUTHORS
TITLE
JOURNAL

COMMENT
Center : Baylor College of Medicine
Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: HANK
Center clone name: RP11-307L1
----- Summary Statistics -----
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy; 18% of reads
Chemistry: Dye-terminator Big Dye; 82% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 186630 bases at least Q40
Consensus quality: 186659 bases at least Q30
Consensus quality: 186660 bases at least Q20
Estimated insert size: 187779; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source
Location=contig of 186660 bp in length.
1.186660:
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /cclone="RP11-307L1"

BASE COUNT
57089 a 37346 c 35962 g 56263 t

ORIGIN

Query Match 49.9%; Score 689; DB 2; Length 186660;
Best Local Similarity 80.1%; Pred. No.2,3e-190;
Matches 850; Conservative 0; Mismatches 175; Indels 36; Gaps 2;

OY 352 gaatacttggtgaatgaagaagcaaatatttggcctataaaagaagaacattctaagcacga 411
||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71843 GAACCTTTAACTTGAGAGACAATGATTGGGTATCGTCGCCGAAGAATAATTTCTAAGTAGCA 71902

OY 412 aagcatcaagaagtactttggtygtcttaaaaggcatccagttcataaaggagcgaca 471
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71903 AAGCATTAAGAAGGTCATCAGATCGTTAAAGGTATTTCAGTTATTAAGAAGGCAGCA 71962

OY 472 gcataaagtcagaaaaatttgcaccctgcacaabtgtagtaaanaaaaagaaaacccaatttc 531
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71963 GCATTAACGCTTTGGAAAAATTGCGACCCGACACTGTGATAGAAAAGAAAACCACATTTTC 72022

QY	532	tgaggggaaatcaagctgctgcagaaatttgcatatgtaatgaagagctgaagttaa	591
Db	72023	TGAGGGAAATTTGAAGCTGTGCTGCACAAATTTGGCATTAAGTAAACAAGAGCGCTAATTTAA	72082
OY	592	tcccaagacacatggggaaanaatctccctggacatgataagaggtcttcaoaagatccat	651
Db	72083	TCCCCAAGACAAATGGGGAAATATCTCCCAAGCCATATGAGAGGTCTTCAATGGCAGCCCAT	72142
OY	652	caaatcactgycctggagagccttagagaaatcgtgtttgttagacaaagccagatgtccc	711
Db	72143	CCCATCACAGGCCCGAGGCGCTAGAGAAATGGTTTGTGAAGGCCAGGCCAGAGGCC	72202
OY	712	gtgctgtgtgcagcctagaaactggttgcctgtgt	747
Db	72203	ATGCTGTGTGAGCTCTAGAGCACTTGTGTGCTCCCTGTATACCGAGCTGCTGCAGTGGTGGCT	72262
OY	748	-----ccagtaattcagctgtggtgttagaggggtgcgaagccccaagccttgg	796
Db	72263	GAAGGGGCGCAGCATAGAGCTTGGGCGGTGGTGGAGGGGTGGAACCCCAAGCCCTTGG	72322
OY	797	cagcttccaagtgtgtgttgccctgtgggtgcgaagaagctcaagaatltgagtttggaa	856
Db	72323	CAATTTCATGTGTGTGTGTGAGCTTAGGGGTGCACAGAAAGTAAATTTGAGCCTTGGGAG	72382
OY	857	cctccaatcagatttcaagaagatataatggaaaaccccttgatgcccagcgagaaattgtct	916
Db	72383	CCTCCACCTTGATTTTCAGAGAGATGTATGGAAATGCTCGATGTCCAGSAGAAAGTTTGCT	72442
OY	917	gttaggggtggggttccctcaatggagaacctctgcgaaggtatgatacaaaaggaatatgttgg	976
Db	72443	TCAAGGGGTGGGGCCCTCATATGGAGAACTCTGTCTAGGGCATGTCAAGAGGGAAATGTGGGG	72502
OY	977	tgggaagcccccacaagaagatccccagltgggggtctcatatlaaagcttggaaagaagt	1036
Db	72503	TTGGAGGCCCTAACAACAGAGTCCCTACTGTGGGACCATATTAAGGGAGCTGTGAGAAGGGG	72562
OY	1037	ccaccatctctccagactccagaaggttaatctcaactgcagctgcagcatgtgacctgaa	1096
Db	72563	CCACCATCTCTCCAGACCCCGAATGCTATGATCCACCAACACTGTGCACATGTATCCCTGGA	72622
OY	1097	aaatccacagacactcaagtgcagcctgtlgaagcagcagagatgtgagttctgtacctac	1156
Db	72623	AAACCCACAGAGACTCAACACAGCCCATGAAGAGCAGCAAGAGGAGGAGCCTGTATACCTCG	72682
OY	1157	aaaacgttagtgcagagcttgacccaagaccgtgggaatcattccctctgcatgttcatgac	1216
Db	72683	AAAGCCACAAACAAAGAGCTTCCCAAGACCAATGGGAAACCCACTCTTGATCATAGTGTGAC	72742
OY	1217	ctgaagcttgagacatgagatgttcaaaaagagatcattttagagctttaaagattgactcccc	1276
Db	72743	CTAATGTGAAACATGTCAGTCCAAAGAGATCTAATTAGAACTTTGAGATTGTGACTACCT	72802
OY	1277	aclygaatttcgactataltggggccgtga-ccccctgttlttgccaatttlttccatt	1335
Db	72803	TCTGGTTTCAGACTTCGATAGGGGCAATATAGGCCCTTGTGTTGGCAATTTTCCCATTT	72862
OY	1336	tgggaactgcgcgatattaccaaatgccttaactccattgtga	1376
Db	72863	TAGAACAGCTGTATTTACTCAATGCTGTATGCGCATTTGCA	72903

RESULT	10		
LOCUS	AC091982/C		
DEFINITION	AC091982	170368 bp	DNA
ACCESSION	AC091982		linear
VERSION	AC091982.3		PR1 30-AUG-2001
KEYWORDS	HTG.	GI:15375177	
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 170368)		

1 (bases 1 to 170368)

AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 170368)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 170368)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Aug 30, 2001 this sequence version replaced gi:14579765. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu Quality: Phrap Quality >=40 99.8% of Sequence; Estimated Total Number of Errors is 0.2. STS Content: SHGC-10855 G13672 SHGC-53350 G36743 WT-14854 G22310 SHGC-64132 G38710. Location/Qualifiers 1. 170368 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="RP11-54C4" BASE COUNT 50024 a 37145 c 34373 g 48826 t ORIGIN
Query Match	49.7%; Score 686.6; DB 9; Length 170368;
Best Local Similarity	81.9%; Pred. No. 1.le-189;
Matches 830; Conservative	0; Mismatches 149; Indels 34; Gaps 2
403	taagcagcaagcacttaagaggctgagctggctgctcttaagaagcattcagttcataag 462
25774	TAAAGCAACAAAGCACTTAAGACAATTATGTGGTGTCTGTAAAGCACTCGGTTTATTAAG 25715
463	gagagcagagataaagagctcagaagaatttgcacccctgacacatgctatataaagaagaaa 522
25714	GGAAGAGAGCACTAAATTTGGAAATTTGGAGCGCTGACAAAGTGCTCTGAAGAAAGAAAAA 26555
523	cccattctcagaggagaaattcaagctcgtgcgcagaaatttcatactgtaatlgagagct 582
25654	CCCATTTTCTGTGGAGAAATTGGAAGCCACAGCTGAAGAAATTTGCATTAAGTAAGAGAGGCC 25595
583	gaatttcatctctaagaacaaatgggaaaataatctcctgcgacatgfcagaggcttcaaca 642
25594	AAATGTATATCTCCCAACAAACATGGGAAAAATGTCCTCGGAGCAATGTCAGAGGTTCTTCACA 25535
643	gcagtcatacaatcacatcagtcctgcgagcgcctcaggaagaaaatggtcttgcgagacagagcc 702
25534	GCAAGCACTCCCATCAACAGGCCCAAAAGCGTTTGGAGAAATCGTTTCGTGGGCGCAAGGCC 25475
703	agggtccctgtcgtctgtgcagcctcagagacttgcgcctgtctccagatthaattcaagc 762
25474	AGGGTCCCTGTGTGTGTGGGCTGTGAAGACTGGTGCCTCTGATCCACGACATCCAGC 25415
763	-----tgctgcctcaagaggtgcgaagcccca 789
25414	CATGACTAAAGAGTGCCAAAGTACATTCACAGTGGTTCTTTCAGAAAGGGAGAGGCCCA 25355
790	gccttcgagacttcacagctgtgttgaagccctgtgcgacaaagaatcagaatcgaagt 849
25354	GCCTTGGCAGCTTCCACAGTGGCAATTGAGCCTGCGGTTGCACAGAACTCAAGAAATTGAAGT 25295
850	ttggagaaactcaatcagatttcataagaagataatatgaaacccctgtatgcccagagcagaa 909

Db	25294	TTGGTAACCTCTGCCTTACATTTTCAAGAAGTGATGAAACCCCTGTGATGCCAACGAANA	25235
Qy	910	gtttctctctaggggtcgttctctatggaaacccctctgaaggtgatgcataaaggaa	969
Db	25234	GTCTGCTCAGAGGGGTGGGGCCCTCATGGAGAACCCTCTGTAGGGAGTGCACAAAGGAAA	25175
Qy	970	tgtttggtggagagcccccacacagagtcgccagcttgaggctccactagtagagctgtg	1029
Db	25174	TGTGGGGTTGAAGACCCCCACACAGAGTCCCTACTTGTGGGACCTCTAGTGGAGCTGTG	25115
Qy	1030	aagaagtcaccatctctccagaactccagaaggtagatccactgaacgtcagaatgt	1089
Db	25114	AAGAGGGCCACCACTCCAGACCCCAAGANTGTGATCACCATTCACACTGTGA	25055
Qy	1090	gacctaaatccacagacactcagcttgccagccttgaagagcagagtagagctgt	1149
Db	25054	GCGTGGAAAAGCCACAGACACTCAACGCTAGCCTTGTAAAGACACAGAGGAGGCTT	24995
Qy	1150	accctacaaaacccgtagtgtgcagaagctgcaccaagaacgctlggaaatctacctctcat	1209
Db	24994	ACCCTGCAAAAGCCACAGAGGGCATACTGCTGCCAAGACATGGAGACCCACCTTGATCA	24935
Qy	1210	tcatgacctgagcagagaaactatgagctaaagaagatatatttggagcttaagattga	1269
Db	24934	GTGTACCTGCTATGATGAGACGAGAGTCAAAAGACATCATTTTGGATCTTTAAGACTTGA	24875
Qy	1270	ctgcccacacttgattctcgtgactatactatggggccgtgta-ccccctgttcttgccaat	1328
Db	24874	CTGTCTACTGTGATTTACAGACTCGCATGGGCGCTGTAGCCCTTTGTTTGGCCAAATGTC	24815
Qy	1329	tctcatcttgggaactgcggtatttaccacaatgcctgtacctcatgtatgtg	1381
Db	24814	TCCCATTTAAGATGGCTGTATTATTCATACCTATCCACCCCATGTATCTGAG	24762
RESULT 11			
AC010312		67984 bp DNA linear	PRI 23-AUG-2001
LOCUS	AC010312		
DEFINITION	Homo sapiens chromosome 5 clone CTB-62P13, complete sequence.		
ACCESSION	AC010312		
VERSION	AC010312.5		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Direct Submission		
REFERENCE	2 (bases 1 to 67984)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint		
REFERENCE	3 (bases 1 to 67984)		
AUTHORS	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell		
AUTHORS	Drive, Walnut Creek, CA 94598, USA		
TITLE	4 (bases 1 to 67984)		
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell		
TITLE	Drive, Walnut Creek, CA 94598, USA		
JOURNAL	5 (bases 1 to 67984)		
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.		
AUTHORS	Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell		
TITLE	Drive, Walnut Creek, CA 94598, USA		
JOURNAL	On Aug 23, 2001 this sequence version replaced gi:11079410.		
COMMENT	Draft Sequence Produced by DOE Joint Genome Institute		
	www.jgi.doe.gov		

Finishing Completed at Stanford Human Genome Center									
www.shgc.stanford.edu									
Quality: Phrap Quality >=40 99.7% of Sequence;									
Estimated Total Number of Errors is 0.1.									
STS Content:									
SHGC-10855 G13672.									
Location/Qualifiers									
1. 67984									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/chromosome="5"									
/clone="CTB-62P13"									
BASE COUNT 19168 a 13195 c 14662 g 20959 t									
ORIGIN									
Query Match	49.7%;	Score 686;	DB 9;	Length 67984;					
Best Local Similarity	81.9%;	Pred. No. 1.6e-189;							
Matches 830;	Conservative 0;	Mismatches 150;	Indels 34;	Gaps					
402	ctaagcagcaagacatctcaagagagcgtgacgtggtgtctgttaaggcatcagttccataa	461							
7119	CTAAGCAGCAAGACATCTCAAGAAGTATGTGCTGTCTGTTAAAGCACCCTTTTATATA	7178							
462	gggagcgagcagcataagagttcagaanaatttcacccctgcacacatgltgataaaaaa	521							
7179	GGGAGCGAGCAGCATTAATAATTTGGAAATTTTCACCCCTGCATCATGTGCTAGAAAAA	7238							
522	accacatttctbaggagggaatcaatcacgcctgycgcaggaatctgcatabgtataggagc	581							
7239	ACCCATTTTCTAGAGAAATTAAGGCCAGCTGAAAGAAATTTTGCAATGATATGAGGAGC	7298							
582	tgaagttaatccataagaacaatggggaataatctcctgcagacatgctcagagcttcac	641							
7299	CAAAATGTTAATCCCAAGACATATGGGGAAATATGCTCCGGACATGTGAGAGCTTTAC	7358							
642	agcagtcacatcaatcaatcactgyccttggaggcctaggaagaatcgttttggagacagcc	701							
7359	AGCACCACACTCCCAATCACAGGCCCAAAGGCTTTGGAGAAATCGTTGGTGGCCAAAGC	7418							
702	cagggtccctgctgctgtctgacagctagagacttggtgcacctgtgtcccaagttaattcag	761							
7419	CAGGCTCCCTGTGCTGTGTGCGGTCTAGAGATTTGTGTGCTGTCTCCAGCATTCCAG	7478							
762	C-----tgagcttcagaggggtgcagcccca	788							
7479	CCATGATCTAATAAAGTGCCAAAGTACATATTCAGGTGTGTTCTTCAAGAGTGGAACCCCA	7538							
789	agccttgcagcttccaaagtgtgttgcagccctgttgggtgcagaagatcaagaattcag	848							
7539	AGCCTTGGCAGCTTCCACGTGTCATTGACCTTGCAGGCGCCACAGAAATTCAGATTGAAG	7598							
849	tttgggaacctcacaatcagatttcagaagatalatlgaaaccccttgcattgcacagcaga	908							
7599	TTTGGTAACTCTGCTGCCTAGATTTCAGAAAGTGTATGAGAAACCCCTGATGCCCAAGCAA	7658							
909	agtttcgtgtgaggggtgtgggttctctatagtagaactcttcgcaaggttagtacaagaaggaa	968							
7659	AGTCTGCTGCAGAGGGGTGGGGCCCTTATGAGAAACCTCTGCTAGGGCAGTGTGCAAGAGGGA	7718							
969	atgttggttggagagccccaacagagatcccccagtgagggttccatctagttaagagcttga	1028							
7719	ATGTGGGGTTGAAGCCCCCACACAGATTCCTACTTTGGGCACCTCTCTAGTGAAGCTTGTA	7778							
1029	gaagaagtcacccaatcctcctcagacttcagaaaggtatagatccactgcagcttcagcaga	1088							
7779	GAAAGCGCCACACACATCTCCAGAGCCCGAATATGATATTCACATGCAAACTTGCACTGTG	7838							
1089	tgcttgaataaatccacagacatctagttgcagcctgtgtaaagcagcagagatagtgatctg	1148							
7839	AGCCTTGGAAGAGCCACAACAACCTCAACCTTACGCTTATGAAAGCACCCAGAGGGAGGCTC	7898							
1149	taccctacaaaacccttagtgcagagcctgcacaaagacgltggaaatctacactcttgac	1208							

Db	7899	TACCGTGAAGCCACAGGGGCGATAGCTCCCAAGACCATTGGGAACCCCTTCGATC	7958
Qy	1209	gtcatgaccctgggacgctggaacatgtgaatcacaagagatcatttggagccttaagtattg	1268
Db	7959	AGTGTGACCTGGCAATATGAAACAGCTGGAGTCACAAAGAGATCATTTTGGATCTTTTAAGACTTGG	8018
Qy	1269	actgcccacatgtgatttcggacttataatgagggcccgta-ccccttcttcttggccaattt	1327
Db	8019	ACTGTCTACTATGGATTTCAGACTCCGCAATAGGGGCGCTGTAGCCCCCTTTTGTGGCCAAATG	8078
Qy	1328	tttccatttgggaactgcgctgatttaccacaatgcctgtacctccatctgtatgtatg	1381
Db	8079	CTCCCATTTTGGATGATGGCTGTATTTATTTCCCAATACCTGTACCCCTTTTACTCTAG	8132
RESULT 12			
AC091005/c			
LOCUS	AC091005	168502 bp	DNA linear PRI 11-DEC-2001
DEFINITION	Homo sapiens chromosome 15, clone RP11-100821, complete sequence.		
ACCESSION	AC091005		
VERSION	AC091005.9	GI:17488655	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 168502)		
JOURNAL	2 (bases 1 to 168502)		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 168502)		
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,		
	Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,		
	Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,		
	Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K.,		
	Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D.,		
	Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,		
	Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W.,		
	Iliev,I., Johnson,R., Jones,C., Karatas,A., Lacomque,K.,		
	Lamares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,		
	Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,		
	McEwan,P., McKernan,K., Mcpheeters,R., Meldrim,J., Meneus,L.,		
	Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,		
	Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,		
	Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,		
	Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P., Roman,J.,		
	Rosetti,M., Royce,A., Santos,R., Schauer,S., Schubbach,R., Seaman,S.,		
	Sevojny,P., Sougez,C., Spencer,B., Stange-Thomann,N.,		
	Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,		
	Theodore,J., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,		
	viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,		
	Zainou,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome		
REFERENCE	Research, 320 Charles Street, Cambridge, MA 02141, USA		
AUTHORS	3 (bases 1 to 168502)		
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,		
	Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,		
	Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,		
	Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,		
	Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,		
	Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,		
	Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,		
	Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,		
	Jones,C., Kamt,A., Karatas,A., Kells,C., Lacomque,K.,		
	Lamares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,		
	Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,		
	McCarthy,M., McKernan,P., McKernan,K., Mcpheeters,R., Meldrim,J.,		
	Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,		
	Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,		
	Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,		
	Raymond,C., Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,		

TITLE	JOURNAL	REFERENCE	AUTHORS
Direct Submission			
Submitted (01-SEP-2001)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	4 (bases 1 to 168502)	
Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barua,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Deavellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Laroque,K., Lamacaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenaga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhang,P., Pierre,N., Pollara,V., Raymond,C., Relta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tasfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
Genome Center			
Center: Whitehead Institute/ MIT Center for Genome Research			
Center code: MIBR			
Web site: http://www-seq.wi.mit.edu			
Contact: sequence_submissions@genome.wi.mit.edu			
Project Information			
Center project name: LI0835			
Center clone name: 1008_C_21			
Location/Qualifiers			
1. 168502			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="15"			
/map="15"			
/clone="RP11-1008C21"			
/clone.lib="RPC1-11 Human Male BAC"			
/rpt_family="12"			
140. 191			
/rpt_family="12"			
194. 218			
/rpt_family="(CA)n"			
complement(222. 381)			
/rpt_family="FRAM"			
384. 464			
/rpt_family="12"			
412. 417			
/note="<30 qual SNGL region"			
667. 672			
/note="<30 qual SNGL region"			
799. 803			
/note="<30 qual SNGL region"			
1130. 1158			
/rpt_family="(CA)n"			
complement(1886. 1985)			

unsure	/rpt_family="MIR" /note=<30 qual SNGL region "
repeat_region	3489. .3530 /rpt_family="A)n"
repeat_region	4319. .4348 /rpt_family="AT_rich"
repeat_region	4775. .4903 /rpt_family="MER85"
repeat_region	5137. .5175 /rpt_family="AT_rich"
repeat_region	5198. .5557 /rpt_family="MLT2F"
repeat_region	5558. .5595 /rpt_family="CA)n"
repeat_region	5596. .5624 /rpt_family="MLT2F"
repeat_region	complement(6521. .6587) /rpt_family="MIR"
unsure	6377. .6382 /note=<30 qual SNGL region "
repeat_region	complement(6414. .6621) /rpt_family="MIR"
repeat_region	6744. .6963 /rpt_family="L2"
repeat_region	complement(7718. .8106) /rpt_family="MLT1B"
repeat_region	8610. .8732 /rpt_family="L2"
repeat_region	complement(9591. .9711) /rpt_family="MIR"
repeat_region	complement(10164. .10479) /rpt_family="AluB"
repeat_region	complement(11119. .11206) /rpt_family="MIR"
repeat_region	13093. .13155 /rpt_family="CGTG)n"
repeat_region	13446. .13751 /rpt_family="AluSx"
repeat_region	14102. .14669 /rpt_family="MLT1E2"
repeat_region	14671. .14973 /rpt_family="AluB"
repeat_region	15003. .15117 /rpt_family="GAA)n"
repeat_region	complement(15227. .15559) /rpt_family="AluB"
repeat_region	15738. .16076 /rpt_family="MLT1L"
repeat_region	17512. .17576 /rpt_family="MLT1L"
repeat_region	17763. .17980 /rpt_family="MER58A"
repeat_region	18048. .18230 /rpt_family="MER5A"
repeat_region	18427. .18562 /rpt_family="MIR3"
repeat_region	complement(18660. .18828) /rpt_family="MER3B"
repeat_region	complement(18937. .19116) /rpt_family="MER5A"
repeat_region	complement(19619. .20141) /rpt_family="L2"
repeat_region	complement(20145. .20670) /rpt_family="MER69"
repeat_region	complement(20669. .21182) /rpt_family="MER69"
repeat_region	complement(21183. .21315) /rpt_family="L2"
repeat_region	complement(21860. .22086) /rpt_family="MER58A"
repeat_region	complement(22428. .22915)

5 10 15 20 25 30 35 40
Phrap Value Range

Version: 1.01 gxf.
Location/Qualifiers
Source

1.73390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-547c5"
19.106
/rpt_family="MIR"
complement(826.1067)
/rpt_family="MSTB"
complement(1070.2183)
/rpt_family="MSTB-internal"
complement(2211.2493)
/rpt_family="MSTB-internal"
complement(2494.2875)
/rpt_family="MSTB"
complement(3718.4776)
/rpt_family="L2"
5066.5137
/rpt_family="MERSA"
complement(5082.5228)
/rpt_family="MERSA"
5802.6088
/rpt_family="MLT2"
6169.6327
/rpt_family="MLT2H"
complement(6751.6903)
/rpt_family="MLT2H"
6930.7009
/rpt_family="HERVL"
8972.9004
/rpt_family="(CA)n"
complement(9689.9781)
/rpt_family="L2"
complement(9782.10068)
/rpt_family="AlusX"
complement(10069.11228)
/rpt_family="L2"
complement(13516.14031)
/rpt_family="L1PA14"
14261.14435
/rpt_family="MIR"
14857.15267
/rpt_family="MLT1B"
15776.15819
/rpt_family="(CA)n"
16497.16657
/rpt_family="MIR"
16714.17041
/rpt_family="MLT1F1"
complement(17251.17555)
/rpt_family="AlusP"
18449.18514
/rpt_family="GA-rich"

Query Match 49.48; Score 682.4; DB 9; Length 73390;
Best Local Similarity 80.2%; Pred. No. 1.8e-188;
Matches 853; Conservative 0; Mismatches 176; Indels 35; Gaps 3;

OY 352 gaataattgctaatgaggaacaatttggctaatagaagaagaatttcaagcaga 411
|||||
DB 65115 GAACATTTCAGCTTGAGAGATGATTTGCGCTATCTCGCGAAGAAATTTCTAAGAAACA 65056
OY 412 aagactcaagaggtgactggtcgtttaaggcattcaggttcataaaggagagcaga 471
|||||
DB 65055 AACGATTTCAGAGGTGACTGCTGCTGTATAGCATTAAAGTTTATTAAGGAGAGAGA 64996

OY 472 gcaataaggttcagaataattgcacccctgacaaatgtatataaagaagaacccctttc 531
|||||
DB 64995 GCATAAAGATTCAGAAATTTGCAACCTGACATGTGTAGTAAAGAAACCCCTTTGCG 64936
OY 532 tgaagggaattcaagcctgagcagaataattgcataatgtaatgagagcgaatgttaa 591
|||||
DB 64935 TGAGGAGAAATTTCAAGCCAGCTGCAAGAAATTTGCTAAGTTAAACAGAGCCAAATTTAA 64876
OY 592 tcccaagaacaattggggaataatccctgagacatgtaacaggtcttcacagcagttccat 651
|||||
DB 64875 TTGCCAAGACATAGGGGAAATGTCTCCAGGGCAGTGCAGAGGTTCTCAGACGACCCCT 64816
OY 652 caaatcactgctgagcagcctagagaagaatgtttgtgagagagcccaaggtccct 711
|||||
DB 64815 CCCATCAGAGGCTCAGAACCTGAGAGGAATGTTTCATGGCCGACGCCAGGTCCCT 64756
OY 712 gtgctgtgtaagccttagaagactgtgcccgtgtcccaatlaatta----- 760
DB 64755 GTGCTGTGTGACGCTTAGAGACTGTGCTGCTGTCCAGCTGTTCCAGCCATGGTTAA 64656
OY 761 -----gctgtgctcagaagggtcgaagccccaagccttgga 798
DB 64695 AAGGGGCCAATGTAGAGCTCAGAGCCATGGCTTCAAGGGTGCAGACTCCAGCCTTGGA 64636
OY 799 gctccaaagtgtgttgagcctgtggtgcaagaagtcgaagaatltgaggttggaacc 858
|||||
DB 64635 GCTTCACATGATGTGCAGAGCCCTGCAAGTGCACAGAAATCAAGAAATCAAGGTTTGGGAAC 64576
OY 859 tccaatcagatttcagaagatatatgaaacccctgagtgcccaagcagaagttgtgcgt 918
|||||
DB 64575 TCTGCTGATGATTTCAGAAATATATGGAACACCTGATGCCAGCAGAAATTTGTCTGC 64516
OY 919 aggggtggtgtcccatgtaagaacctctgcaaggtaagtaacaagggaatgttgggt 978
DB 64515 AGGGGTGGGGGCTACCGGAAACCTCTGCTGGGCAATGTGGAAGGAAATGTGTGT 64456
OY 979 ggaagccccaacaagaagtcaccaagtgagggtcccatctagtagagctgtgagaagaatcc 1038
|||||
DB 64455 GAAGCCCCCAACAGATGCTCTACTAGGGGAGTGCCTGATGGAATGTGAGAAAGAGGCC 64396
OY 1039 accatccccaagactccaagaaggtgagatccatctgacagctctgacagctgtgcgaaga 1098
|||||
DB 64395 ACCGCTCTCCAGGC-CCGAATGTGATGCCATGACAGCTGCGACCATTTCCCTGAAAA 64337
OY 1099 atcccaagaacctcagtgccagcctgtgaaagcagaagatgagatgtgtgacacaa 1158
|||||
DB 64336 AGCCACACACTTATATGCCAGCCCTTGAAGSCACTGAGAGAGGCTGTGCTGCA 64277
OY 1159 aaccgtagtgacagagctgacccaagaacgtgggaatcactcctctgcaatgtcatgacct 1218
|||||
DB 64276 AGCCACAGGATGAGAGCTGTGCCAAGACCATGGAACCCAGCTGTATGACATGACACT 64217
OY 1219 ggaagtgagacatgagcacaagaagatcttggagccttaagattgagctgagccac 1278
|||||
DB 64216 GGATGTGAGACCTGAGTCAAGAGATCTTTGGAACCTTTAAAGTTGAGAGCCAC 64157
OY 1279 tgaatttggaactataggggcccgt-accocctgttctggccaatttttccattg 1337
|||||
DB 64156 TGCAATTTTGACATGCAAGGCGCCCTGTAAACCCCTTGTGTTGTCCAAATTTCCCATTTG 64097
OY 1338 gaactgcccatttaccacaatgctgtacccctccattgtatgtag 1381
|||||
DB 64096 GAATGCTGTATTTATCAATATACCTGTAACCCCATGTATGTAG 64053

RESULT 14
AC079614 131215 bp DNA linear PRI 09-JAN-2002
LOCUS
DEFINITION Homo sapiens BAC clone RP11-42414 from 2, complete sequence.
AC079614
ACCESSION AC079614.7 GI:15431230
VERSION
KEYWORDS
SOURCE human.

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
HP-20.
GN HP-20.
OS Tamias sibiricus (Siberian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Tamias.
OX NCBI_TaxID=64680;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2148336; PubMed=11602349;
RA Ono M., Hosoe Y., Azuma S., Shoji M., Nara K., Kondo N., Shiba T.,
RA Takamatsu N.;
RT "HNF-1 regulates the liver-specific transcription of the chipmunk HP-
RT 20 gene.";
RL Gene 277:121-127(2001).
DR EMBL; AB067779; BAB68362.1; -
SQ SEQUENCE 196 AA; 21330 MW; B07D17EBFA94D2FD CRC64;

Query Match 13.0%; Score 324; DB 11; Length 196;
Best Local Similarity 46.3%; Pred. No. 1.9e-21;
Matches 68; Conservative 21; Mismatches 46; Indels 12; Gaps 4;
QY 1 IVIVPLITAVIEHVAAPP---AHPRPPEVGP-----PGAPGLPYTGSEISEMTKC 51
DB 7 LAFLVLMVNLNDQYSCGPGPYGPGVPGVPGPRGPPGQPCAGRPDPGPKGPGSVKC 66
QY 52 PCGDIERSAFYTKLSGKPLPFPKPIITGVLYNQRDLKAMGVFCGRPGNYTSFDVE 111
DB 67 PCF--ERSAFYTKLSGKPLPFPPESEVPTFEVLYNQRDLKASTGVNCFVGRNTHFSFVE 124
QY 112 LHCKYNIWLMRKQI-LANKEEISKOQ 137
DB 125 LVHCKYIGLMKNHIOVMEKHOLSKNE 151

RESULT 3
Q9H720 PRELIMINARY; PRT; 222 AA.
AC Q9H720;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL 24.1 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiyasu S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024120; BAB14833.1; -
SQ SEQUENCE 222 AA; 24079 MW; 64EED514F9F541D9 CRC64;

Query Match 11.0%; Score 275; DB 4; Length 222;
Best Local Similarity 40.0%; Pred. No. 6.1e-17;
Matches 76; Conservative 22; Mismatches 64; Indels 28; Gaps 7;
QY 223 RPR---RWFECGTGPGS--LCVQP-----RDIVPCVPVNSAVASEGASPRWQLPSGV 271
DB 26 RPRGLGKDDFGALGSPALCRLQWTWPAFQKQPC--LKEAKVQLGF---WLQREYV 79
QY 272 EPYGA-----KSRIVWEPPIRFOKIYGNPMMPQKFAVGVGSSWRTSARVQK 321

DB 80 QGIGSFHGLSLQGHRSGLRLGNLHLDFFKRCMEKMPDDQAEVCCIG-GPSWTSARAAMK 138
QY 322 GNVGMEPPHVRPSGAPSSRAVRRSPSSRLQKGRSTDSLOHPEKSTDPQCPVNAAG 381
DB 139 GNVGLEPPHRRPPTVALPSGAVRRGPLSSRPLNGRSTDSLHCAPERAADTQCPMAAGRE 198
QY 382 SVPYKTVAAE 391
DB 199 AIRCKTTGAD 208

RESULT 4
Q921K4 PRELIMINARY; PRT; 295 AA.
ID Q921K4
AC Q921K4;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLLAGEN ALPHA 1 TYPE X (FRAGMENT).
GN COL10A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20310874; PubMed=10853827;
RA Marks S.A., Lundmark C., Christerson C., Wurtz T., Odgren P.R.,
RA Sellert M.F., Mackay C.A., Mason-Savay A., Popoff S.E.;
RT "Endochondral bone formation in toothless (postepetrotic) rats:
RT failures of chondrocyte patterning and type x collagen expression.";
RL Int. J. Dev. Biol. 44:309-316(2000).
DR EMBL; AJ131848; CAA10518.1; -
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; Clq.
DR Pfam; PF01391; Collagen.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q.
DR PROSITE; PS0113; Clq.
FT NON_TER 1
SQ SEQUENCE 295 AA; 30012 MW; FF43B1548028813E CRC64;

Query Match 6.8%; Score 169; DB 11; Length 295;
Best Local Similarity 28.4%; Pred. No. 3.7e-07;
Matches 54; Conservative 19; Mismatches 69; Indels 48; Gaps 6;

QY 18 AGPPAHPRPPEVGPAGPLPYTGSEISEMTKCPCPD----- 55
DB 101 AGPPGPRGRTGHTGRLPGPPGPPGPPGQAVIPDGFTRSGQRPRLSGMPLVSANOG 160
QY 56 ---IERSAFYTKLSKPLPFPKPIITGVLYNQRDLKAMGVFCGRPGNYTSFDVE 112
DB 161 VTGMPVSAFTYLSAAVPAVGAPIPFDEILYNROQHYDPRSGIFTCIKIPGLYFSYTHV 220
QY 113 H--HCKVNIWLMRKQILANKEEISK-----QOSIOEVT-----VWLLKAFSFTREAEH 158
DB 221 KGHVWVGLKXNGPTWYTYDEYSKGLDQASSATIMELTENDQWVLQPL-----NSES 274
QY 159 K---SSENLH 165
DB 275 NGLYSSEYVH 284

RESULT 5
Q96D07 PRELIMINARY; PRT; 744 AA.
ID Q96D07;
AC Q96D07;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHECTICAL 73.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC013581; AAI13581.1;
KM Hypothetical protein.
SQ SEQUENCE 744 AA; 73364 MW; 2BC1B0955DEC9A3 CRC64;

Query Match 6.5%; Score 162.5; DB 4; Length 744;
Best Local Similarity 30.4%; Pred. No. 5.1e-06;
Matches 42; Conservative 17; Mismatches 50; Indels 29; Gaps 5;

OY 19 GPPAHPRPPEEYVGPAGL-----PYTGEI-----SEMTKCPDPI 56
DB 554 GPPGLPGRPPGPPGPPVMPPTPPGGEYLPMGLGIDGVKPPHAYGAKKNGGPAY 613
OY 57 ERSATTVLSGKLPDPFKPIITGVLYNAQRDLKAMGVACRVPNGYSSPDELHCK 116
DB 614 EMPATTAELTAPFPVPGAPVKENKLLYNGRONYNPQTGIFTCVGVYFAVAVH---HCK 670
OY 117 -VNIMLRKQILANKKEI 133
DB 671 GGNVAV---ALFKNNEPV 685

RESULT 6
O96NP0 PRELIMINARY; PRT: 216 AA.
AC O96NP0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ30449 FIS, CLONE BRACE2009274.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM;
RA Nitomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto Y., Isono Y.,
Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
Suzuki Y., Sugano S., Nagahari K., Mshino Y., Nagai K., Isogai T.,
NEDO human cDNA sequencing project.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK055011; BAB70839.1;
SQ SEQUENCE 216 AA; 23140 MW; 976A49B029A84D36 CRC64;

Query Match 6.3%; Score 158; DB 4; Length 216;
Best Local Similarity 28.1%; Pred. No. 2.4e-06;
Matches 59; Conservative 19; Mismatches 68; Indels 64; Gaps 7;

OY 209 CORSSOOSTKSLAMP-----RRKMPG--TGPGSLCV-----QPRDLVPCVPYNSAV 255
DB 15 CORSSW---LPLOYRPPGRLKNGKMGFQSSQSPANCIILGTWCPASOPIQLQMLMLKGAN 71
OY 256 ASEGASPKPMQLPSGVEPVGAKKSRLEWV-----EPPI 288
DB 72 VOLG-----PWLDORVATSLGFGFHVGLMVKRKORRRFGSLHLDPRGCMEMPGCGRRPPL 127

OY 289 RPOKIYGNPMPMPKRAVGVSSWRTSARVQKGNVMEPPHRRVPSGAPSSAARRSPS 348
DB 128 QGQSPHGDPLQO-----CRGNGVLEPSQSRVSTRALPNCAGVGRGPS 170
OY 349 SRLQGRSTDSQHPPEKSTDPQCOPIKAA 378
DB 171 SRPHKGSSTDSLHCVPKAVGTCCOPIKAA 200

RESULT 7
O9D2V4 PRELIMINARY; PRT: 744 AA.
AC O9D2V4:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROCOLLAGEN, TYPE VIII, ALPHA 1.
GN COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamaneke I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boilelli D., Bojunga N., Carninci P., de Bona M.F.,
Brownstein W.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilmink L.,
Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuhl S.,
Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DR EMBL: AK018742; BAB31383.1;
DR MGD: MGI:88463; Col8a1.
DR InterPro: IPR001073; Clq.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; Clq; 1.
DR Pfam: PF01391; Collagen; 7.
DR PRINTS: PRO0007; COMPLEMENTC1Q.
DR SMART: SM00110; Clq; 1.
DR PROSITE: PS01113; Clq; 1.
SQ SEQUENCE 744 AA; 73581 MW; C659BDCBCBDEB9C CRC64;

Query Match 6.3%; Score 157.5; DB 11; Length 744;
Best Local Similarity 32.0%; Pred. No. 1.4e-05;
Matches 40; Conservative 14; Mismatches 46; Indels 25; Gaps 4;

OY 19 GPPAHPRPPEEYVGPAGL-----PYTGEI-----SEMTKCPDPI 57
DB 555 GPPGLPGRPPGPPGPPVMPPTPPGGEYLPMGLGIDGVKPPHAYGAKKNGGPAY 614
OY 58 RSAFTVLSGKLPDPFKPIITGVLYNAQRDLKAMGVACRVPNGYSSPDELHCK- 116
DB 615 MPATTAELTAPFPVPGAPVKFKLLYNGRONYNPQTGIFTCVGVYFAVAVH---HCKG 671
OY 117 VNIMLRKQILANKKEI 121

Db 672 GNVWV 676

RESULT 8

0921S8 PRELIMINARY; PRT: 744 AA.

ID 0921S8 PRELIMINARY; PRT: 744 AA.

AC 0921S8; 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE PROCOLLAGEN, TYPE VIII, ALPHA 1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC011061; AA011061.1; -.

KM Collagen.

SO SEQUENCE 744 AA; 73621 MW; BF7A7ED79D8463AA CRC64;

Query Match 6.3%; Score 157.5; DB 11; Length 744;

Best Local Similarity 32.0%; Pred. No. 1.4e-05;

Matches 40; Conservative 14; Mismatches 46; Indels 25; Gaps 4;

QY 19 GPPAHRPEVEGPGAPGL---PQYTG-----ISEMTKPCPDIE 57

Db 555 GQPGLEGGPPGPPGPPPPVMTSPGGEFLPDMGLGIDGKPPHAYAGKKGHGEPAYE 614

QY 58 RSAPFYVLSGKLPKPKPIFTGVLYNARDLKEMGVACRVPGNYSSFDVLEHCK- 116

Db 615 MPAFTELVPEPPVGPAYKFKDLKLYNGRONVNPOTGIFTCVGVYFAYHV---HCKG 671

QY 117 VNIML 121

Db 672 GNVWV 676

RESULT 9

09N178 PRELIMINARY; PRT: 675 AA.

ID 09N178 PRELIMINARY; PRT: 675 AA.

AC 09N178; 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE TYPE X COLLAGEN.

GN COL10A1.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-21015405; PubMed-11130976;

RA Nielsen V.H., Bendixen C., Arnbjerg J., Sorensen C.M., Jensen H.E.,

RA Shukri N.M., Thomsen B.

RT "Abnormal growth plate function in pigs carrying a dominant mutation

RT in type X collagen."

RL Mamm. Genome 11:1087-1092(2000).

DR EMBL; AF222861; AA03271.1; -.

DR InterPro; IPR001073; Clq.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF00386; Clq; 1.

DR Pfam; PF01391; Collagen; 6.

DR PRINTS; PRO0007; COMPLEMENTC1Q.

DR SMART; SM00110; C1Q; 1.

DR PROSITE; PS01113; C1Q; 1.

KM Collagen.

SO SEQUENCE 675 AA; 65447 MW; 26397B10310383F9 CRC64;

Query Match 6.3%; Score 157; DB 6; Length 675;

Best Local Similarity 32.1%; Pred. No. 1.4e-05;

Matches 45; Conservative 14; Mismatches 53; Indels 28; Gaps 5;

QY 19 GPPAHRP-----PEEVGPGAPGLPQYTG-----GEISE-----MTKPCPDIE 57

Db 491 GPPGPPGPKHAGEPGLPGPPGPGGAVPPEGFVKEGAFVSANQVGTMPV---- 546

QY 58 RSAPFYVLSGKLPKPKPIFTGVLYNARDLKEMGVACRVPGNYSSFDVLEH--HC 115

Db 547 -SAPFYVLSKAPPAIGAPLPFDKILNCGOHYDPKTLFTCRIPGIFYFSYHIVKGTAA 605

QY 116 KVNIMLRKQIILANKKEISK 135

Db 606 WYGLYKNGPVPVMTYDEYVK 625

RESULT 10

095J95 PRELIMINARY; PRT: 194 AA.

ID 095J95 PRELIMINARY; PRT: 194 AA.

AC 095J95; 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE ADIPONECTIN (FRAGMENT).

GN ADP1.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-ADIPOSE TISSUE;

RA Kabir M., Anantharayan S., Ionut V., Kim S.P., Van Citters G.W.,

RA Dea M.K., Bergman R.N.

RT "Regulation of Adiponectin gene expression in the fat-fed dog."

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF417206; AA09702.1; -.

FT NON_TER 1 194

FT NON_TER 1 194

SO SEQUENCE 194 AA; 20890 MW; 3AA3D947D187AF9A CRC64;

Query Match 6.2%; Score 155.5; DB 6; Length 194;

Best Local Similarity 34.5%; Pred. No. 3.5e-06;

Matches 41; Conservative 16; Mismatches 55; Indels 7; Gaps 2;

QY 31 GPGAPGLPQYTGISEMTKCPDIERSAPFYVLSGKLPKPKPIFTGVLYNARDLK 90

Db 67 GPPGFPETPGKGEPEESAV----VHSAPSVGLESHITVPNPIRFTKIFYNLQNHVD 121

QY 91 EANGVACRVPGNYSSF--DVLEHCKVNIMLRKQIILANKKEISKQSGIOEVTWLL 147

Db 122 GTTGKFRHCNIPGLYFSYHITVLYLNDVKSILYKKDKALFTFYDQYQKKNDAQSGSVLL 180

RESULT 11

095M04 PRELIMINARY; PRT: 240 AA.

ID 095M04 PRELIMINARY; PRT: 240 AA.

AC 095M04; 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE ADIPOSE TISSUE-SPECIFIC PROTEIN ADIPO Q.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecorida; Bovidae;

OC Bovidae; Bovinae; Bos.

NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-21369933; PubMed-11382781;

RA Sato C., Yasukawa Z., Honda N., Matsuda T., Kitajima K.;
 RT "Identification and Adipocyte Differentiation-dependent Expression of
 RT the Unique Distal Acid Residue in an Adipose Tissue-specific
 RT Glycoprotein, Adipo O";
 RL J. Biol. Chem. 276:28849-28856(2001).
 DR EMBL; AF269230; AAK58902.1; -.
 SQ SEQUENCE 240 AA; 26091 MW; C6253BA803B9A668 CRC64;

Query Match 6.2%; Score 155.5; DB 6; Length 240;
 Best Local Similarity 29.7%; Pred. No. 4.7e-06;
 Matches 43; Conservative 27; Mismatches 64; Indels 11; Gaps 3;

QY 31 GPPGAPGLPYOTGEISEMTKCPDIERSAFYVLSGKLPPIIFGVLVNAORDLK 90
 DB 85 GPRGPRGTPGRKGEPEALAY-----VYRSASFVGLERTVNVNPIRFTKITYNOQNHVD 139
 QY 91 EMAGVFACRVPGNYYSSDVELH--HCKVNIWLMRKQILANKKEISKOOSIDVTVWLK 148
 DB 140 GSTGKFCYNIPCLYFSYHITVYMKDVKSLFKDKKAVLFYDQYQEKNVDAQSVLLH 199
 QY 149 A----FSFIREAEHKSSENLPDNY 169
 DB 200 LEVGDQWMLQYVEGENHNGVTADNV 224

RESULT 12

ID 090MB3 PRELIMINARY; PRT; 350 AA.

AC 090MB3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MYELOID-SPECIFIC PEROXIDASE (FRAGMENT).
 GN MPX.
 OS Brachydanio rerio (zebrafish) (zebra danio).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 CC Cypriniformes; Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA Lleschke G.J., Oates A.C., Crowhurst M.O., Ward A.C., Layton J.E.;
 RT "Macrophages in embryonic and adult zebrafish."
 RL Blood 0:0-0(2001).
 DR EMBL; AF378825; AAK91662.1; -.
 KW Peroxidase.
 FT NON_TER 1
 SQ SEQUENCE 350 AA; 38647 MW; 4E1EADDA0573F70 CRC64;

Query Match 6.0%; Score 148.5; DB 13; Length 350;
 Best Local Similarity 33.3%; Pred. No. 3.4e-05;
 Matches 34; Conservative 12; Mismatches 33; Indels 23; Gaps 3;

QY 17 VAGPPAHPRPEE-----VGPAPGLP---QYTGISEMTKCPDIERSAFYVLS 66
 DB 244 IPGPRGPRGTPGRKGEPEALAY-----VYRSASFVGLERTVNVNPIRFTKITYNOQNHVD 139
 QY 67 GKLPLEPKPIIFTGVLVNAORDLKEMAGVFACRVPGNYYSSP 108
 DB 291 SILPATAKVVVGVLYNGONHNYNOTSGMFLCOIPGVYEEFF 332

RESULT 13

ID 09BXJ2 PRELIMINARY; PRT; 289 AA.

AC 09BXJ2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE COMPLEMENT-C1Q TUMOR NECROSIS FACTOR-RELATED PROTEIN.
 GN CTRP7.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Piddington C.S., Sheppard P.O., Bishop P., Lasser G.W.;
 RT "Homo sapiens complement-c1q tumor necrosis factor-related protein."
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF329839; AAK17963.1; -.
 DR InterPro; IPR001073; C1q.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF00386; C1q; 1.
 DR Pfam; PF01391; Collagen; 1.
 DR PRINTS; PR00007; COMPLEMENTC1Q.
 DR SMART; SM00110; C1Q; 1.
 DR PROSITE; PS01113; C1Q; 1.
 SQ SEQUENCE 289 AA; 30683 MW; A61609FF86D26946 CRC64;

Query Match 5.8%; Score 144; DB 4; Length 289;
 Best Local Similarity 36.2%; Pred. No. 6.7e-05;
 Matches 38; Conservative 11; Mismatches 46; Indels 10; Gaps 2;

QY 19 GPPAHPRPEEYGPAPGLPYOTGEISEMTKCPDIE-RSAFTVLSGKLPPII 77
 DB 116 GPPIPGPRGDRGEGDPPPLPGV-----CRGSIYLSKSAFSGITTSTPEERPII 166
 QY 78 FTGVLVNAORDLKEMAGVFACRVPGNYYSSDVELHCKVNIWLM 122
 DB 167 FKKVLFENGEHNVNATGKFTICAFPGIYFSYDITLANKHLAIGLV 211

RESULT 14

ID 095JD7 PRELIMINARY; PRT; 243 AA.

AC 095JD7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ADIPONECTIN.
 GN APM1.
 OS Macaca mulatta (rhesus macaque).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 CC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADIPOSE TISSUE;
 RX MEDLINE=21232234; PubMed=11334417;
 RA Hotta K., Funahashi T., Bodkin N.L., Ortmeier H.K., Arita Y.,
 RA Hansen B.C., Matsuzawa Y.;
 RT "Circulating concentrations of the adipocyte protein adiponectin are
 RT decreased in parallel with reduced insulin sensitivity during the
 RT progression to type 2 diabetes in rhesus monkeys."
 RL Diabetes 50:1126-1133(2001).
 DR EMBL; AF404407; AAK92202.1; -.
 SQ SEQUENCE 243 AA; 26264 MW; 49A45DAF2B4613FD CRC64;

Query Match 5.6%; Score 140.5; DB 6; Length 243;
 Best Local Similarity 28.8%; Pred. No. 0.00011;
 Matches 40; Conservative 17; Mismatches 51; Indels 31; Gaps 3;

QY 17 VAGPPAHP-----RPPEYGPAPGLPYOTGEISE--MTCPCPD----- 55
 DB 39 MAGIPGHGHNCGVRDGRDGRGKGEKGDPLGPKDGTGETGTAGEGRGRFPGLG 98
 QY 56 -----IRSAFTVLSGKLPPIIFTGVLVNAORDLKEMAGVFACRVPGNYY 105

Db 99 RKSEPEGAGYVRSFAFSGLETVYTVENMDIFRTKIFYNQNNHHDGSGTKFHCNIPGLY 158

QY 106 SDFVELHHCXVNIWLMR 124

159 FATHITVIMKDYKVSLEFR 177

Db

RESULT	15			
062789				
ID	062789	PRELIMINARY:	PRT:	173 AA.
AC	062789:			
DT	01-AUG-1998 (TEMBLrel. 07, Created)			
DT	01-AUG-1998 (TEMBLrel. 07, Last sequence update)			
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)			
DE	COLLAGEN VIII (FRAGMENT).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=9825;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	TISSUE-AORTIC SMOOTH MUSCLE;			
RA	Reichenberg S., Plenz G., Robenek H.;			
RL	Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF054891; AAC08999.1; -			
DR	InterPro: IPR001073; Clq.			
DR	Pfam: PF00386; Clq. 1.			
DR	PRINTS: PRD0007; COMPLENNTClQ.			
DR	SMART: SM00110; ClQ. 1.			
DR	PROSITE: PS01113; ClQ. 1.			
FT	NON_TER	1		
FT	NON_TER	173	173	
SO	SEQUENCE	173 AA;	18973 MM;	03489B6FEAAB6CAD CRC64;

Query Match	5.5%;	Score 137;	DB 6;	Length 173;
Best Local Similarity	30.3%;	Pred. No. 0.00014;		
Matches	37;	Conservative	18;	Mismatches 45;
			Indels	22;
			Gaps	5

[illegible]

```
Search completed: June 30, 2002, 11:54:29
Job time: 145 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:52:24 ; Search time 13.41 Seconds

(without alignments)
1325.300 Million cell updates/sec

Title: US-09-997-610-2

Perfect score: 2494
Sequence: 1 IVVIVPLITRAVIEHVEVAP.....GQFFPGTAVTQCLYLHGM 459

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwisProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	330	13.2	196	HP20_TAMSI
2	221	8.9	215	HP25_TAMSI
3	171	6.9	743	CA18_MOUSE
4	167.5	6.7	215	HP27_TAMSI
5	165.5	6.6	674	CA1A_CHICK
6	163.5	6.6	744	CA18_HUMAN
7	163.5	6.6	744	CA18_RABIT
8	159.5	6.4	680	CA1A_MOUSE
9	159	6.4	674	CA1A_BOVIN
10	156	6.3	680	CA1A_HUMAN
11	152.5	6.1	247	APM1_MOUSE
12	148.5	6.0	244	APM1_HUMAN
13	146.5	5.9	635	CA28_HUMAN
14	132.5	5.3	245	C10C_HUMAN
15	124.5	5.0	258	C18F_HUMAN
16	118.5	4.8	245	YPC2_ECOLI
17	115.5	4.6	258	C18F_MOUSE
18	111.5	4.5	1035	TAC2_MOUSE
19	111	4.5	255	GLIC_MOUSE
20	110.5	4.4	170	CA28_MOUSE
21	110.5	4.4	251	C10B_HUMAN
22	110	4.4	3210	CENF_HUMAN
23	107.5	4.3	670	DVL1_HUMAN
24	105.5	4.2	245	C10A_MOUSE
25	105	4.2	670	DVL1_MOUSE
26	103	4.1	633	XRC1_HUMAN
27	101.5	4.1	551	GRF3_YEAST
28	101	4.0	253	C10B_MOUSE
29	101	4.0	1026	TAC2_HUMAN
30	100	4.0	245	C10A_HUMAN
31	96.5	3.9	1005	BLU2_HUMAN
32	96.5	3.9	1356	V195_HUMAN
33	96	3.8	246	C10C_MOUSE

34	95.5	3.8	763	1	FXM1_HUMAN	Q08050 h forkhead
35	95	3.8	1618	1	NEST_HUMAN	P48681 homo sapien
36	94.5	3.8	600	1	SP96_DICDI	P14328 dictyostell
37	94.5	3.8	628	1	V7OK_TYMC	P28478 turnip yell
38	94	3.8	695	1	DVL1_MOUSE	P51141 mus musculu
39	94	3.8	695	1	DVL1_RAT	O9wrb9 rattus norv
40	94	3.8	698	1	YB06_YEAST	P38283 saccharomyc
41	94	3.8	806	1	B1MA_EMENT	P17885 emericeila
42	93.5	3.7	1237	1	B3A2_RABIT	P48746 oryctolagus
43	93	3.7	1162	1	LEPR_MOUSE	P48356 mus musculu
44	93	3.7	3149	1	TECU_EBV	P03186 epstein-bar
45	92.5	3.7	253	1	C10B_RAT	P31721 rattus norv

ALIGNMENTS

```

RESULT 1
ID HP20_TAMSI STANDARD: PRT; 196 AA.
AC Q06575;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hibernation-associated plasma protein Hp-20 precursor (Hibernator-
DE specific blood complex, 20 kDa subunit).
OS Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogathi; Scuridae; Scurinae;
OC Tamias.
OX NCBI_TaxID=64680;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93180798; PubMed=8441393;
RA Takematsu N., Ohba K., Kondo J., Kondo N., Shiba T.;
RT "Hibernation-associated gene regulation of plasma proteins with a
RT collagen-like domain in mammalian hibernators.";
RL Mol. Cell. Biol. 13:1516-1521(1993).
RN [2]
RN SEQUENCE OF 24-58; 66-99; 104-129; 132-136 AND 150-184.
RP TISSUE=Plasma;
RX MEDLINE=92112696; PubMed=1730610;
RA Kondo N., Kondo J.;
RT "Identification of novel blood proteins specific for mammalian
RT hibernation.";
RL J. Biol. Chem. 267:473-478(1992).
CC -!- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A
CC 140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE
CC HIBERNATION SPECIFIC.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -!- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE
CC PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION
CC CEASES.
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D12974; BA002351.1;
CC InterPro: IPR001073; C1Q.
CC InterPro: IPR000087; Collagen.
CC Pfam: PF00386; C1Q; 1.
CC PRINTS: PR00007; COMPLEMENTC1Q.
CC SMART: SM00110; C1Q; 1.
CC PROSITE: PS01113; C1Q; 1.

```


CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE OF 48-674 FROM N.A., AND SEQUENCE OF 103-117 AND 453-466.
RA	MEDLINE=66168227; PubMed=3082876;
RX	Ninomiya Y., Gordon M., van der Rest M., Schmid T., Linsemayer T.,
RA	Olsen B.R.;
RT	"The developmentally regulated type X collagen gene contains a long
RT	open reading frame without introns."
RL	J. Biol. Chem. 261:5041-5050(1986).
RN	[2]
RP	SEQUENCE OF 1-75 FROM N.A.
RX	MEDLINE=89054019; PubMed=2461368;
RA	Luvalle P., Ninomiya Y., Rosenblum N.D., Olsen B.R.;
RT	"The type X collagen gene. Intron sequences split the 5'-untranslated
RT	region and separate the coding regions for the non-collagenous amino-
RL	terminal and triple-helical domains."
RN	J. Biol. Chem. 263:18378-18385(1988).
RN	[3]
RP	REVISIONS TO C-TERMINUS.
RA	MEDLINE=89380199; PubMed=2476437;
RY	Yamauchi N., Benya P.D., van der Rest M., Ninomiya Y.;
RT	"The cloning and sequencing of alpha 1(VIII) collagen cDNAs
RT	demonstrate that type VIII collagen is a short chain collagen and
RT	contains triple-helical and carboxyl-terminal non-triple helical
RT	domains similar to those of type X collagen."
RL	J. Biol. Chem. 264:16022-16029(1989).
CC	-1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPIC
CC	CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC	MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC	-1- SUBUNIT: HOMOTRIMER.
CC	-1- PWM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC	UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	-1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC	-1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@sib-sib.ch).
CC	-----
CC	EMBL; M13496; AAA48736.1; ALT_SEQ.
DR	EMBL; J04194; AAA48634.1; -. .
DR	PIR; A31896; A31896.
DR	InterPro; IPR001073; C1q.
DR	InterPro; IPR000087; Collagen.
DR	Pfam; PF00386; C1q; 1.
DR	Pfam; PF01391; Collagen; 6.
DR	PRINTS; PR00007; COMPLEMENTC1Q.
DR	SMART; SM00110; C1Q; 1.
DR	PROSITE; PS01113; C1Q; 1.
RW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW	Cartilage; Collagen; Signal.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	MOD_RES
FT	MOD_RES
FT	SEQUENCE
Q0	674 AA; 66434 MW; EMB48B1EF17AB145 CRC64;

ID	SEQUENCE	STANDARD	PROT	744 AA
Db	494 SGEGLPEPPGPPGPGGOSTIPBQYVKESRSLSSMSMKRGAQGLATMPV-----SNF			548
Qy	62 TWKISGLPLPEKPIIFFTGVLYNAORDLKEAMGFACRVPNGVYSSFDVELHCK-VNIM			120
Db	549 TVLISKAVPEATVPKEDKILYNRQOHDPDPTGTFCTCPGLYFYFHV--HAKGTNW			605
Qy	121 L 121			
Db	606 V 606			
RESULT	6			
Ca18_HUMAN	STANDARD:	PROT:	744 AA.	
AC	P27658:			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 1(VIII) chain precursor (Endothelial collagen).			
GN	COL8A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91231001; PubMed=2029894;			
RA	Muraqaki Y., Mattei M.-G., Yamauchi N., Olsen B.R., Nimniya Y.;			
RT	"The complete primary structure of the human alpha 1 (VIII) chain and assignment of its gene (COL8A1) to chromosome 3.";			
RL	Eur. J. Biochem. 197;615-622(1991).			
CC	-1- FUNCTION: MAJOR COMPONENT OF THE DESCMET'S MEMBRANE (BASEMENT MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.			
CC	-1- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION WITH ALPHA 2(VIII) TYPE COLLAGENS.			
CC	-1- PRIM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.			
CC	-1- MISCELLANEOUS: FOUR CONSECUTIVE GLY-PRO-PRO TRIPLETS ARE PRESENT AT THE C-TERMINUS OF THE TRIPLE-HELICAL REGION. THESE MAY PROVIDE			
CC	-1- THE HIGH THERMAL STABILITY OF THIS REGION.			
CC	-1- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.			
CC	-1- SIMILARITY: CONTAINS 1 C10 DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).			
CC	-----			
DR	EMBL; X57527; CAA0748.1; -.			
DR	PIR; S15435; S15435.			
DR	MIM; 120251; -.			
DR	InterPro; IPR001073; C1Q.			
DR	InterPro; IPR000087; Collagen.			
DR	Pfam; PF00386; C1Q; 1.			
DR	Pfam; PF01391; Collagen; 7.			
DR	PRINTS; PRO0007; COMPLEMENTC1Q.			
DR	SMART; SM00110; C1Q; 1.			
DR	SMART; SM00110; C1Q; 1.			
DR	PROSITE; PS01113; C1Q; 1.			
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;			
KW	Glycoprotein; Cell adhesion; Collagen; Signal.			
FT	SIGNAL	1	28	
FT	CHAIN	29	744	COLLAGEN ALPHA 1(VIII) CHAIN.
FT	DOMAIN	29	117	NONHELICAL REGION (NC2).
FT	DOMAIN	118	571	TRIPLE-HELICAL REGION (COL1).
FT	DOMAIN	572	744	NONHELICAL REGION (NC1).
FT	DOMAIN	609	744	C1Q.
QO	SEQUENCE	744 AA;	73443 MW;	7852C22582B377A3 CRC64;


```
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; X67348; CAA47763.1; -.
DR EMBL; X65121; CAA46237.1; -.
DR EMBL; X63013; CAA44741.1; -.
DR EMBL; Z21610; CAA79736.1; -.
DR PIR; S28807; S28807.
DR PIR; S31216; S31216.
DR PIR; S22215; S22215.
DR MGD; MGI:88445; Col10a1.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COMPLEMNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal.
FT SIGNAL 1 18
FT CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 680 NONHELICAL REGION (NC1).
FT DOMAIN 545 680 C1Q.
FT CONFLICT 248 248 P -> L (IN REF. 3).
FT CONFLICT 286 286 A -> S (IN REF. 2).
FT CONFLICT 306 306 L -> F (IN REF. 3).
FT CONFLICT 417 417 T -> S (IN REF. 3).
FT CONFLICT 451 451 R -> K (IN REF. 4).
FT CONFLICT 500 500 H -> L (IN REF. 3).
FT CONFLICT 567 567 A -> C (IN REF. 3).
FT CONFLICT 569 569 I -> H (IN REF. 3).
FT CONFLICT 571 572 FD -> Y (IN REF. 3).
FT CONFLICT 635 635 Q -> T (IN REF. 3).
SQ SEQUENCE 680 AA; 66775 MW; FE984CA99AF708E2 CRC64;

Query Match 6.4%; Score 159.5; DB 1; Length 680;
Best Local Similarity 29.6%; Pred. No. 0.00012;
Matches 55; Conservative 18; Mismatches 68; Indels 45; Gaps 8;

QY 19 GPPAHPRP-----PEVGPPEAGELPOLY-----TGEISEMTKPCPCDIER----- 58
DB 490 GPPGPPGPGHSGEPGLGPPGPPGPPGPOAVMPDGFYKAOQRRLSGMPLVSANHGVTGM 549
QY 59 --SAFTVKLSGKLPLPEPPIIFTGVLYNAORDLKEAMGVACRVPGNYSSPDVELH--H 114
DB 550 PYSAFTVILSKAPVAGAPVAPPEDELILYKROOHYPRSGIFTCKIPGIIYFSYVHVAKGTH 609
QY 115 CKVNIWLMKROILANKKEISK-----QOSIOEVT-----VWLKAFSFIKEAERK--- 159
DB 610 VVWGLKNGTPTMYTYDEYSKGYLDQAGSAIMELTENDQVWLOLP-----NAESNGLY 663
QY 160 SSENLA 165
DB 664 SSEYVH 669

RESULT 9
```

```
CA1A_BOVIN
ID CA1A_BOVIN STANDARD; PRT; 674 AA.
AC P23206;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=91113131; PubMed=1703407;
RA Thomas J.T., Kwan A.P.L., Grant M.E., Boot-Handford R.P.;
RT Isolation of cDNAs encoding the complete sequence of bovine type X
RT collagen. Evidence for the condensed nature of mammalian type X
RT collagen genes.
RL Biochem. J. 273:141-148(1991).
CC -1- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; X53556; CAA37624.1; -.
DR PIR; S13301; S13301.
DR InterPro; IPR001073; C1q.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COMPLEMNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cartilage; Collagen; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 674 COLLAGEN ALPHA 1(X) CHAIN.
FT DOMAIN 19 56 NONHELICAL REGION (NC2).
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.
FT DOMAIN 520 674 NONHELICAL REGION (NC1).
FT DOMAIN 539 674 C1Q.
FT DISULFID 194 197 BY SIMILARITY.
FT MOD_RES 460 460 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 463 463 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. - -) (POTENTIAL).
SQ SEQUENCE 674 AA; 65546 MW; CD4CA73A03E004CA CRC64;

Query Match 6.4%; Score 159; DB 1; Length 674;
Best Local Similarity 31.8%; Pred. No. 0.00013;
Matches 41; Conservative 16; Mismatches 56; Indels 16; Gaps 3;

QY 18 ACPPAHPRPPEVGPPEAGELPO-----YGTGEISEMTKPCPCDIRSAFTYKLSK 68
DB 501 ACPGGLPPEPPGPPGPPGVALLPEDFVAKGRPEVSANQGTGMPV-----SAFTYILSKA 555
QY 69 LPLPEKPIIFTGVLYNAORDLKEAMGVACRVPGNYSSPDVELH--HCKVNIWLMKROKI 126
```


DB 556 YPAIGPDPDKLVNKQOHYDPRGIFTCIKIPGIYFSTHHVKGCTHANVGLYKNGTPV 615
 QY 127 LANKREISK 135
 : : I
 DB 616 MYTYDEYIK 624
 RESULT 10
 ID CAIA_HUMAN STANDARD: PRT: 680 AA.
 AC 003692;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(X) chain precursor.
 GN COL10A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92109659; PubMed=1764025;
 RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
 RA Solomon E., Grant M.E., Boot-Handford R.P.;
 RT "The human collagen X gene. Complete primary translated sequence and
 RT chromosomal localization.";
 RL Biochem. J. 280:617-623(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93012005; PubMed=1397333;
 RA Reichenberger E., Beyer F., Luvaille P., Olsen B.R., von der Mark K.,
 RA Bertling W.M.;
 RT "Genomic organization and full-length cDNA sequence of human collagen
 RT X.";
 RL FEBS Lett. 311:305-310(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Beyer F., Lamm M.B., von der Mark K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Williams S.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 52-680 FROM N.A.
 RX MEDLINE=92267014; PubMed=1587271;
 RA Apte S.S., Seidlin M.F., Hayashi M., Olsen B.R.;
 RT "Cloning of the human and mouse type X collagen genes and mapping of
 RT the mouse type X collagen gene to chromosome 10.";
 RL Eur. J. Biochem. 206:217-224(1992).
 RN [6]
 RP SEQUENCE OF 561-666 FROM N.A.
 RX MEDLINE=91243838; PubMed=2037056;
 RA Apte S., Mattei M.-G., Olsen B.R.;
 RT "Cloning of human alpha 1(X) collagen DNA and localization of the
 RT COL10A1 gene to the q21-q22 region of human chromosome 6.";
 RL FEBS Lett. 282:393-396(1991).
 RN [7]
 RP SEQUENCE OF 547-655 FROM N.A.
 RX MEDLINE=92077285; PubMed=1743401;
 RA Reichenberger E., Aigner T., von der Mark K., Stoeck H., Bertling W.;
 RT "In situ hybridization studies on the expression of type X collagen
 RT in fetal human cartilage.";
 RL Dev. Biol. 148:562-572(1991).
 RN [8]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97253959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).

RN [9]
 RP VARIANTS SMCD ASP-598 AND PRO-614.
 RX MEDLINE=94136476; PubMed=8304336;
 RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
 RA Grant M.E., Boot-Handford R.P.;
 RT "Amino acid substitutions of conserved residues in the
 RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
 RT occur in two unrelated families with metaphyseal chondrodysplasia
 RT type Schmid.";
 RL Am. J. Hum. Genet. 54:169-178(1994).
 RN [10]
 RP VARIANT SMCD ARG-591.
 RX MEDLINE=94272470; PubMed=8004099;
 RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
 RT "Additional mutations of type X collagen confirm COL10A1 as the
 RT Schmid metaphyseal chondrodysplasia locus.";
 RL Hum. Mol. Genet. 3:303-307(1994).
 RN [11]
 RP VARIANT SMCD VAL-618.
 RX MEDLINE=95181449; PubMed=7876225;
 RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
 RT "Type X collagen multimer assembly in vitro is prevented by a Gly618
 RT to Val mutation in the alpha 1(X) NCI domain resulting in Schmid
 RT metaphyseal chondrodysplasia.";
 RL J. Biol. Chem. 270:4558-4562(1995).
 RN [12]
 RP VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
 RX MEDLINE=95331767; PubMed=7607655;
 RA Bonaventure J., Chamnade F., Maroteaux P.;
 RT "Mutations in three subdomains of the carboxy-terminal region of
 RT collagen type X account for most of the Schmid metaphyseal
 RT dysplasias.";
 RL Hum. Genet. 96:58-64(1995).
 RN [13]
 RP VARIANT SMCD PRO-600.
 RX MEDLINE=96375754; PubMed=8782043;
 RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
 RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
 RT "Mutations within the gene encoding the alpha 1 (X) chain of type X
 RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
 RT not several other forms of metaphyseal chondrodysplasia.";
 RL J. Med. Genet. 33:450-457(1996).
 RN [14]
 RP VARIANTS SMCD GLU-18 AND ARG-18.
 RX MEDLINE=97220591; PubMed=9067753;
 RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
 RT "Mutations in the N-terminal globular domain of the type X collagen
 RT gene (COL10A1) in patients with Schmid metaphyseal
 RT chondrodysplasia.";
 RL Hum. Mutat. 9:131-135(1997).
 RN [15]
 RP VARIANTS SMD GLU-595.
 RX MEDLINE=99057503; PubMed=9837818;
 RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
 RA Nakamura Y.;
 RT "Mutation of the type X collagen gene 'COL10A1' causes
 RT spondylometaphyseal dysplasia.";
 RL Am. J. Hum. Genet. 63:1659-1662(1998).
 RN [16]
 RP VARIANT SMCD CYS-597.
 RX MEDLINE=99069781; PubMed=9852679;
 RA Sawai H., Ida A., Nakata Y., Koyama K.;
 RT "Novel missense mutation resulting in the substitution of tyrosine by
 RT cysteine at codon 597 of the type X collagen gene associated with
 RT Schmid metaphyseal chondrodysplasia.";
 RL J. Hum. Genet. 43:259-261(1998).
 CC -I- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTROPHIC
 CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
 CC MINERALIZATION ZONES OF HYALINE CARTILAGE.
 CC -I- SUBUNIT: HOMOTRIMER.
 CC -I- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE

SECRETED INTO SERUM.
-1- INDUCTION: DURING HORMONE-INDUCED ADIPOSE DIFFERENTIATION AND ACTIVATED BY INSULIN.
-1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
-1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U37222; AAA80543.1; -
DR EMBL: U49915; AAB06706.1; -
DR EMBL: AF044466; AAK13417.1; -
DR EMBL: AK001138; BAB22597.1; -
DR MGD: MGI:106675; ACRP30.
DR InterPro: IPR001073; C1q.
DR Pfam: PF00386; C1q; 1.
DR Pfam: PF01391; Collagen; 1.
DR PRINTS: PR00007; COMPLEMENTC1Q.
DR SMART: SM0110; C1Q; 1.
DR PROSITE: PS01113; C1Q; 1.
KW Collagen; Signal; Repeat; Hydroxylation; Plasma; Polymorphism.
FT SIGNAL 1 17
FT CHAIN 18 247
FT DOMAIN 45 110
FT DISULFID 39 39
FT MOD_RES 47 47
FT MOD_RES 50 50
FT MOD_RES 56 56
FT MOD_RES 65 65
FT MOD_RES 79 79
FT MOD_RES 98 98
FT MOD_RES 107 107
FT MOD_RES 113 113
FT VARIANT 50 50
FT CONFLICT 74 74
FT CONFLICT 117 117
FT CONFLICT 148 148
FT CONFLICT 243 243
SQ SEQUENCE 247 AA; 26841 MW; 137B687D873988C4 CRC64;

Query Match 6.1%; Score 152.5; DB 1; Length 247;
Best Local Similarity 29.7%; Pred. No. 0.00012;
Matches 44; Conservative 23; Mismatches 64; Indels 17; Gaps 4;

QY 31 GPPGAPGLPQYATGEISEMKPCPDLEBSATFYKLSGKLPKPIITFGVLYNAORDLK 90
DB 93 GRRGFPFGTGRGGEPEALY-----MYRSATSGVLETRVTPVPRIFKIFLNQDNHYD 147
QY 91 EAMGFACHPGPNYSSPDELH--HCKYINIMLRKOILANKKEISKOOSIOEVTWLLK 148
DB 148 GSTGKFCYCNIPGLYFYSYHITVYMKDKVSLFKKDKAVLFYTDQYQEKNDVDSGSLVLLH 207
QY 149 A-----FSFIREAEHKSENLHPDNV 169
DB 208 LEVGDQVWLVQYVGDDH---NGLYADNV 232

RESULT 12
APML_HUMAN STANDARD; PRT; 244 AA.
AC 015848;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Adiponectin precursor (30 kDa adipocyte complement-related protein)

DE (ACRP30) (Adipose most abundant gene transcript 1) (apm-1) (gelatin-binding protein).
GN APML OR ACRP30 OR GBP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Adipose tissue;
RX MEDLINE=96224171; PubMed=8619847;
RA Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y., Matsubara K.;
RT "cDNA cloning and expression of a novel adipose specific collagen-like factor, apm1 (Adipose most abundant gene transcript 1).";
RL Biochem. Biophys. Res. Commun. 221:286-289(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99196984; PubMed=10095105;
RA Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M., Nakano Y., Shimizu N., Tomita M.;
RT "Organization of the gene for gelatin-binding protein (GBP28).";
RL gene 229:67-73(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Schaeffler A., Orso E., Paltzsch K.D., Buechler C., Drobnik W., Fuert A., Scholmerich J., Schmitz G.;
RT "The human apm-1, an adipocyte-specific gene linked to the family of TNF's and to genes expressed in activated T cells, is mapped to chromosome 1q21.3-q23, a susceptibility locus identified for familial combined hyperlipidemia (FCH).";
RL Biochem. Biophys. Res. Commun. 260:416-425(1999).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=204417747; PubMed=10961870;
RA Yokota T., Oritani K., Takahashi T., Ishikawa J., Matsuyama A., Uuchi N., Kihara S., Funahashi T., Tenner A.J., Tomiyama Y., Matsuzawa Y.;
RT "Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages.";
RL Blood 96:1723-1732(2000).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20440368; PubMed=10982546;
RA Uuchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H., Hotta K., Nishida M., Takahashi M., Murauchi M., Ohmoto Y., Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.;
RT "Adiponectin, an adipocyte-derived plasma protein, inhibits endothelial NF-kappaB signaling through a camp-dependent pathway.";
RL Circulation 102:1296-1301(2000).
CC -1- FUNCTION: IMPORTANT NEGATIVE REGULATOR IN HEMATOPOIESIS AND IMMUNE SYSTEMS. MAY BE INVOLVED IN ENDING INFLAMMATORY RESPONSES THROUGH ITS INHIBITORY FUNCTIONS. INHIBITS ENDOTHELIAL NF-KAPPAB SIGNALING THROUGH A CAMP-DEPENDENT PATHWAY. INHIBITS TNF-ALPHA-INDUCED EXPRESSION OF ENDOTHELIAL ADHESION MOLECULES.
CC -1- SUBUNIT: HOMODIGOMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED IN PLASMA.
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE ADIPOSE TISSUE.
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D45371; BAA08227.1; -
DR EMBL: AB012165; BAA86716.1; -
DR EMBL: AB012164; BAA86716.1; JOINED.

Search completed: June 30, 2002, 11:54:49
Job time: 145 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:51:29 ; Search time 20.93 Seconds

(without alignments)
2107.263 Million cell updates/sec

Title: US-09-997-610-2

Perfect score: 2494

Sequence: 1 IVVIVPLITAVIEHVEVAGP.....GQFFPGTAVFTOCLYLHCM 439

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	330	13.2	196	2	A48150 hibernation-relate
2	221	8.9	215	2	B48150 hibernation-relate
3	171	6.9	743	1	S23779 collagen alpha 1(V
4	167.5	6.7	215	2	C48150 hibernation-relate
5	165.5	6.6	744	1	S23298 collagen alpha 1(V
6	163.5	6.6	744	1	S23297 collagen alpha 1(V
7	163.5	6.6	744	1	A34246 collagen alpha 1(V
8	159.5	6.4	680	2	S15435 collagen alpha 1(V
9	159.5	6.4	674	2	S13216 collagen alpha 1(V
10	156	6.3	674	2	S13301 collagen alpha 1(V
11	148.5	6.0	680	1	CGH1D collagen alpha 1(X
12	146.5	5.9	635	2	UC4708 collagen alpha 1(X
13	143.5	5.3	635	2	A57131 collagen alpha 2(V
14	132.5	5.3	245	1	C1HUOC collagen alpha 2(V
15	129.5	5.2	1553	2	C1HUOC complement subcomp
16	121.5	4.9	219	2	T14782 rab3 effector prot
17	117.5	4.7	245	2	S12089 hypothetical prote
18	111	4.5	898	2	S69634 hypothetical prote
19	110.5	4.4	170	2	B57131 collagen alpha 2(V
20	110.5	4.4	253	1	C1HUOC complement subcomp
21	110	4.4	813	2	S70795 vasa protein precu
22	110	4.4	1017	2	PC4035 cell-cycle-depende
23	106.5	4.3	1008	2	T49470 phosphatidic acid-
24	105.5	4.2	245	2	S19018 complement subcomp
25	103	4.1	405	4	A61181 homeotic protein H
26	103	4.1	633	2	A36353 DNA repair protein
27	101.5	4.1	551	2	S53377 Gzf3 protein - yea
28	101	4.0	253	2	I49560 complement C1q B c
29	100.5	4.0	1280	2	T00365 hypothetical prote

30	100	4.0	245	1	C1HUQA complement subcomp
31	99.5	4.0	1495	2	T31434 desmin-180 - rat
32	99	4.0	343	2	H64491 hypothetical prote
33	98.5	3.9	992	2	T08772 hypothetical prote
34	97.5	3.9	740	2	S17925 polynucleotide ade
35	97	3.9	207	1	Q08CU2 hypothetical 21.9K
36	97	3.9	355	2	T29932 hypothetical prote
37	96.5	3.9	2187	2	T30826 nascent polypeptid
38	96	3.8	246	2	S29328 complement subcomp
39	96	3.8	620	2	T50150 yeast nrd1-like pr
40	96	3.8	920	2	T43263 cell division prot
41	95.5	3.8	895	2	S74225 leptin receptor, i
42	95	3.8	1618	2	S21424 nestin - human
43	94.5	3.8	481	2	G75253 hypothetical prote
44	94.5	3.8	600	2	S07638 spore coat protein
45	94.5	3.8	628	2	S19150 hypothetical prote

ALIGNMENTS

```
RESULT 1
A48150
hibernation-related protein HP-20 precursor - Siberian chipmunk
C:Species: Eutamias sibiricus (Siberian chipmunk)
C>Date: 16-Feb-1994 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: A48150; C41752
R:Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.
Mol. Cell. Biol. 13, 1516-1521, 1993
A>Title: Hibernation-associated gene regulation of plasma proteins with a collagen-11
A:Reference number: A48150; MUID:93180798
A:Accession: A48150
A:Molecule type: mRNA; protein
A:Residues: 1-196 <TAK>
A:Cross-references: GB:DJ2974; NID:9287467; PIDN:BA02351.1; PID:9287468
A>Note: the source is designated as Tamias asiaticus in Genbank entry TMSHP20A, relea
R:Kondo, N.; Kondo, J.
J. Biol. Chem. 267, 473-478, 1992
A>Title: Identification of novel blood proteins specific for mammalian hibernation.
A:Reference number: A41752; MUID:92112656
A:Accession: C41752
A:Status: preliminary
A:Molecule type: protein
A:Residues: 24-58;66-99;104-129;132-136;137,151-184 <KON>
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal
C:Keywords: glycoprotein; hibernation; plasma
F:1-24/Domains: signal sequence #status predicted <SIG>
F:24-196/Product: hibernation-related protein HP-20 #status experimental <MAT>
F:173-196/Domains: complement C1q carboxyl-terminal homology <C1Q>
F:154/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.2%; Score 330; DB 2; Length 196;
Best Local Similarity 46.9%; Pred. No. 1.6e-18;
Matches 69; Conservative 21; Mismatches 45; Indels 12; Gaps 4;

OY 1 IVVIVPLITAVIEHVEVAGP---AHPRPEEVGP-----PCAPGLPYTGISEMTKC 51
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7 LAIFVLMVNLNDVSCSGPPGPGVGTGVPVGPGRPGQGRACRPPDPRKSGSVAC 66
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 52 PCEDIERSAFTYKLSGRLPLPRKPIIFTGVLVNAQRDLKAMGACRVPNGYSSFDVE 111
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 PCR--ERSAFYKFSGRLLPPESEPVETVLVNTQDLKESGCVFNCVPGNHFVSFDE 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 112 LHHCKYNIWLMKQI-LANKEISKOQ 137
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 LYHCKYKIGLMKNHIOVMEKHQLSKNE 151
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
B48150
hibernation-related protein HP-25 precursor - Siberian chipmunk
```



```

Db      53  MLKCVVRDNAPKA---ACKTPRNAES-----QPHQPLNIF---RCVCPVRKPSSEGGP 101
QY      141 EYTWVLLKAFSPFIREAEHKSSENL-----HPDNV-----IKKNPSEGEKFLA-----E 186
Db      102 ERDMRLHQOFESYKEQVRRKIGGEARRYOGEHKKDDAPTCGICHTKTFADGGHLCSTYCRK 161
QY      187 ICT-CNEELNVNPDNGENISWTCQ--RSSQSI-KSLAMPRRKWFCGTG-----GS 236
Db      162 FCARCGGRVSLRSNNEDKVMVWVNCNLCRKQOEILTKSGA-----WFGSGPQOPSDGT 215
QY      237 LCCVQ-----PROLVPCV-----PVNSAVAS--EGASP-----KPMQLPS 269
Db      216 LSDTATGAGSEVPREKKARLQERSRQTPUSTAIVSSQDTAIFGAPLHRNKGAEPSQAL 275
QY      270 GVEPVGAKKSRIEVMPEPIRFQKIYGNPMMPROKFAVGVSS-WRTSARVYQKGNVWEP 328
Db      276 GPEQKQASRSRS---EPPREKKKA---PGLSEQNGKGGQKSERKRVPKSVVQPG----- 323
QY      329 PHRVPSGAPSSRAVRRSPSSRLQKGRSTDLSLOHVPEK 366
Db      324 -----EGIADEREKREKRETRRLEKGRSQDYSR-PEK 355

```

Search completed: June 30, 2002, 11:53:51
 JOD time: 142 sec

THIS PAGE BLANK (USPTO)

EARLIER APPLICATION NUMBER: 60/056,983
EARLIER FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 6
LENGTH: 236
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-140-804-6

Query Match 7.0%; Score 174; DB 4; Length 236;
Best Local Similarity 31.3%; Pred. No. 3.7e-09;
Matches 47; Conservative 15; Mismatches 74; Indels 14; Gaps 2;

QY 6 VLTAVIEHVEVAGPPAPRPPPEEVEGPPGA-----GDLPOYTGEISEMTKPC 53
DB 48 ILALSLMHVCSNVPGQGPGRMGKPGPGKPGPPGGLPGPPGPMYNCHS 107
QY 54 PDIERSAFTVKSGLPLPFKPIIFTGYLYNAQRDLKEAMGVACRYPGNYTSSFDVELH 113
DB 108 KG--TSAFAVKANLEPPAPSQPVIFKEALHDQGHFDLATGVFTCPGLYQGFHIEAV 165
QY 114 HCKVIMLMRKQIILANKEEISKOOSIOEVT 143
DB 166 QRAVKYSLMRNGTOYMERAREADQGEYHIS 195

RESULT 3
US-08-463-911-3
Sequence 3, Application US/08463911
Patent No. 5869330
GENERAL INFORMATION:
APPLICANT: Scherer, Philipp E.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,911
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH195-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-463-911-3

Query Match 6.7%; Score 167; DB 2; Length 185;
Best Local Similarity 35.2%; Pred. No. 1.2e-08;

Matches 45; Conservative 11; Mismatches 64; Indels 8; Gaps 3;
QY 19 GPPAPRPPPEEVEGPP---GAPGLPOYTGEISEMTKPCPDIERSAFTVKSGLPLPFKP 75
DB 22 GPPGTPGKR---GPPGNVGFGLPGPPGPMYNCHSKG--TSAFAVKANLEPPAPSQP 76
QY 76 IFTGYLYNAQRDLKEAMGVACRYPGNYTSSFDVELHCKVIMLMRKQIILANKEEISK 135
DB 77 VIFKEALHDQGHFDLATGVFTCPGLYQGFHIEAVQRAVKYSLMRNGTOYMERAREAE 136
QY 136 QOSIOEVT 143
DB 137 QDQGEYHIS 144

RESULT 4
US-08-463-911-2
Sequence 2, Application US/08463911
Patent No. 5869330
GENERAL INFORMATION:
APPLICANT: Scherer, Philipp E.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,911
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH195-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-911-2

Query Match 6.1%; Score 152.5; DB 2; Length 247;
Best Local Similarity 29.7%; Pred. No. 5.3e-07;
Matches 44; Conservative 23; Mismatches 64; Indels 17; Gaps 4;

QY 31 GPPGAPGLPOYTGEISEMTKPCPDIERSAFTVKSGLPLPFKPIIFTGYLYNAQRDLK 90
DB 93 GPRGPGTPGKRGEGEREAY----MYRSAFSVGLETRVTVNPVIRFTKIFYNQNNHYD 147
QY 91 EAMGVACRYPGNYTSSFDVELH--HCKVIMLMRKQIILANKEEISKOOSIOEVTWVLK 148
DB 148 GSTGKFCYNIGPLGYFSYHTYWKVDKYSLEFRKRAVLFTYDYOERKNVDAQSGSVLLH 207
QY 149 A-----FSFIREAEHKSSENLPDNY 169
DB 208 LEVGDQVWLQYVGGGDH---NGLYADNV 232

RESULT 5

US-08-463-911-7
 ; Sequence 7, Application US/08463911
 ; Patent No. 5869330

GENERAL INFORMATION:

APPLICANT: Scherer, Philipp E.
 APPLICANT: Lodish, Harvey F.
 TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
 TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS: 7
 ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Millitia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA: US/08/463,911
 APPLICATION NUMBER: US/08/463,911
 FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WH95-05

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 244 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-463-911-7

Query Match 6.0%; Score 148.5; DB 2; Length 244;

Best Local Similarity 30.7%; Pred. No. 1.3e-06;

Matches 43; Conservative 16; Mismatches 48; Indels 33; Gaps 4;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCR-----CPDIE 57

DB 40 MAGIGPHGHNCAFGRDGDTGPEGEKGDPLGPKDICE-TGVPAEGPRGPGIO 98

QY 58 -----RSAFYVLSGKLPFPKPIFTGVLYNAORDLKEAMGVACRYPGNY 104

DB 99 GRKGEPEGAYYRSASFVGLTYTIPMPPIRFTKIFYNQNHDSYGRKHCNIPGLY 158

QY 105 YSFVDELHCKVNIWLMRK 124

DB 159 YFAVHITVYMKDVSLFRK 178

US-09-140-804-3

Sequence 3, Application US/09140804

Patent No. 6197930

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Humes, Jacqueline M.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS

FILE REFERENCE: 97-49

CURRENT APPLICATION NUMBER: US/09/140, 804

CURRENT FILING DATE: 1998-08-26

EARLIER APPLICATION NUMBER: 60/056, 983

EARLIER FILING DATE: 1997-08-26

NUMBER OF SEQ ID NOS: 47
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 3

LENGTH: 244
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-140-804-3

Query Match

Best Local Similarity 6.0%; Score 148.5; DB 4; Length 244;

Matches 43; Conservative 16; Mismatches 48; Indels 33; Gaps 4;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCR-----CPDIE 57

DB 40 MAGIGPHGHNCAFGRDGDTGPEGEKGDPLGPKDICE-TGVPAEGPRGPGIO 98

QY 58 -----RSAFYVLSGKLPFPKPIFTGVLYNAORDLKEAMGVACRYPGNY 104

DB 99 GRKGEPEGAYYRSASFVGLTYTIPMPPIRFTKIFYNQNHDSYGRKHCNIPGLY 158

QY 105 YSFVDELHCKVNIWLMRK 124

DB 159 YFAVHITVYMKDVSLFRK 178

US-09-140-804-8

Sequence 8, Application US/09140804

Patent No. 6197930

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Humes, Jacqueline M.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS

FILE REFERENCE: 97-49

CURRENT APPLICATION NUMBER: US/09/140, 804

CURRENT FILING DATE: 1998-08-26

EARLIER APPLICATION NUMBER: 60/056, 983

EARLIER FILING DATE: 1997-08-26

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 8

LENGTH: 247

TYPE: PRT

ORGANISM: Homo sapiens

US-09-140-804-8

Query Match 5.8%; Score 144.5; DB 4; Length 247;

Best Local Similarity 26.4%; Pred. No. 3.3e-06;

Matches 51; Conservative 25; Mismatches 74; Indels 43; Gaps 6;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGSEISMTCR-----CPDIE 57

DB 43 MAGIGPHGHNCAFGRDGDTGPEGEKGDPLGPKDICE-TGVPAEGPRGPGIO 102

QY 56 -----IERSAFYVLSGKLPFPKPIFTGVLYNAORDLKEAMGVACRYPGNY 105

DB 103 RKGEPEGAYYRSASFVGLTYTIPMPPIRFTKIFYNQNHDSYGRKHCNIPGLY 162

QY 106 SSFVDELH--HCKVNIWLMRKQIILANKKEISKQDSIQEYTWYLKA-----FSFIREA 156

DB 163 FSYHITVYMKDVKSLFKDKAVLFTYDQYDEKKNVDSAGSVLHLLEVQDQWLVQYGDG 222

QY 157 EHKSSENLHPDNV 169

DB 223 DH---NGLYADNV 232

US-09-118-408-3

Sequence 3, Application US/09118408A

Patent No. 6265344

```

; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-30
; CURRENT APPLICATION NUMBER: US/09/118,408A
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/053,154
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-118-408-3

Query Match          5.8%; Score 144.5; DB 4; Length 247;
Best Local Similarity 26.4%; Pred. No. 3.3e-06;
Matches 51; Conservative 25; Mismatches 74; Indels 43; Gaps 6;

QY 17 VAGPPAHP-----RPEEVGPPGAPGLPQYTGISE--MTKCPD----- 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43 MAGIPGHPGNGPGRGDGRDGPGEKSGDAGLGRKGTGVGMGAEGRGPGQTG 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 56 -----IERSAFYKLSGKLPKPIIFGTGLYNAORDLEKAMGVACRPGNTY 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 103 RKGEPEAAWYRSASFVGLERIVVNVPIRFTKIFYNOONHYDSTGKFCYCNIPGLY 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 SPSFVELH--HCKVNIWLMRQILANKKEISKOOSIOEYVWLLKA-----FSPIREA 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 FSHITIVYKDYKVSLEFKDKAVLFTTDQYQEKVNDQASSVLLHLEVGQVWLOYTG 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 157 EHKSENLPDNY 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 DH--NGLYADNY 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-188-930-295
; Sequence 295, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 295
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Rat
US-09-188-930-295

Query Match          5.0%; Score 124; DB 4; Length 243;
Best Local Similarity 31.9%; Pred. No. 0.00034;
Matches 38; Conservative 12; Mismatches 53; Indels 16; Gaps 4;

QY 19 GPPAHPPEEVGPPGAPGLPQYTG-----ISEMKCPCPIERSATYKLS 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 GRDAGAGAGEGEGGRLPGRPEGRGEPGAGPVGAIIPAGECVP--PRSAFSAKRS 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 -GKLPPLF-KPIIFGTGLYNAORDLEKAMGVACRPGNTYSSFDVLEHCKVNIWLMR 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 ESNVPPAPDLPPLPDRVLNLEGGHDAITGKFTCVGVYIFAVHATVIRASLQFDLVK 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 10
US-09-188-930-294
; Sequence 294, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 294
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Rat
US-09-188-930-294
```

```

Query Match          5.0%; Score 123.5; DB 4; Length 294;
Best Local Similarity 31.8%; Pred. No. 0.00051;
Matches 34; Conservative 14; Mismatches 52; Indels 7; Gaps 2;

QY 17 VAGPPAHPPEEVGPPGAPGLPQYTGISEMTKCPDIE-RSAFTYKLSGKLPKPKP 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 VSGTPGKHGIPGKKGPKKKGEPGLP-----PCSGSSRAKSAFSAVATKSPRERLP 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 IIFGTGLYNAORDLEKAMGVACRPGNTYSSFDVLEHCKVNIWLM 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 IKFDKILNMGSHYNASSGKFCVSPGIIYFTYDITLANKHLAIGLV 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 11
US-09-140-804-2
; Sequence 2, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-2
```

```

Query Match          4.9%; Score 121.5; DB 4; Length 243;
Best Local Similarity 27.5%; Pred. No. 0.00059;
Matches 42; Conservative 15; Mismatches 57; Indels 39; Gaps 6;

QY 14 HVEVAGPPAH-----PRPEEVGPPGAPGLPQYTG----- 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 HGLPRTGPHNGSGGLPGRDGRDAGAPAPGEGGGRGGLPGRDGPGRGAGPAG 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 45 -ISEMKCPCPIERSATYKLS-GKLPPLF-KPIIFGTGLYNAORDLEKAMGVACRVP 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 PTPGAGECVP--PRSAFSAKRSRVPSPDAPLPFDRLVNLNLEGGHDAITGKFTCV 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Qy 102 GNYSSFDVLEHCKVNIWLMRKQIIANKEIS 134
| | | | | : : : : : | | :
Db 149 GYYFAVHATVYRASLQFDLVK-----NGESIA 176

RESULT 12
US-09-485-316A-12
; Sequence 12, Application US/09485316A
; Patent No. 6344441
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Poclin, Frances
; TITLE OF INVENTION: Lipoprotein-regulating medicaments
; FILE REFERENCE: GENSET.036APC
; CURRENT APPLICATION NUMBER: US/09/485,316A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT IB98/01256
; PRIOR FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent.pm
; SEQ ID NO 12
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: 1..130
; OTHER INFORMATION: fragment 118..247 of translation from ref Genbank U49915
US-09-485-316A-12

Query Match 4.7%; Score 116; DB 4; Length 130;
Best Local Similarity 27.1%; Pred. No. 0.00079;
Matches 32; Conservative 22; Mismatches 52; Indels 12; Gaps 3;

Qy 61 FTVKLSGKLPPEKPIIFGVLYNAQRDLKAMGVACRVPGNYSSFDVELH--HCKYN 118
| | | | | : : : : : | | : | : : : : : | :
Db 1 FSVGLETRVTPVNPVPIRFKIFYNQONHNDNSGKFCYCNIPGLYFSYHITYVMKDVKS 60

Qy 119 IWLMRQIIANKEISKQOSIOEVTWVLLKA-----FSFIREAHKSENLHPDNV 169
| | | | | : : : : : | | : | : : : : : | :
Db 61 LFKKDKAVLFTYDQYQEKNVDAQSGSVLLHLEVGDDQVWLQVYGDGH---NGLYADNV 115

RESULT 13
US-09-485-316A-13
; Sequence 13, Application US/09485316A
; Patent No. 6344441
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Poclin, Frances
; TITLE OF INVENTION: Lipoprotein-regulating medicaments
; FILE REFERENCE: GENSET.036APC
; CURRENT APPLICATION NUMBER: US/09/485,316A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT IB98/01256
; PRIOR FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent.pm
; SEQ ID NO 13
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Mus musculus

; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: 1..130
; OTHER INFORMATION: fragment 118..267 of translation from ref genbank U37222
US-09-485-316A-13

Query Match 4.6%; Score 114; DB 4; Length 130;
Best Local Similarity 27.1%; Pred. No. 0.0012;
Matches 32; Conservative 22; Mismatches 52; Indels 12; Gaps 3;

Qy 61 FTVKLSGKLPPEKPIIFGVLYNAQRDLKAMGVACRVPGNYSSFDVELH--HCKYN 118
| | | | | : : : : : | | : | : : : : : | :
Db 1 FSVGLETRVTPVNPVPIRFKIFYNQONHNDNSGKFCYCNIPGLYFSYHITYVMKDVKS 60

Qy 119 IWLMRQIIANKEISKQOSIOEVTWVLLKA-----FSFIREAHKSENLHPDNV 169
| | | | | : : : : : | | : | : : : : : | :
Db 61 LFKKDKAVLFTYDQYQEKNVDAQSGSVLLHLEVGDDQVWLQVYGDGH---NGLYADNV 115

RESULT 14
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92132
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 4.4%; Score 110; DB 1; Length 2482;
Best Local Similarity 21.4%; Pred. No. 0.3; Indels 112; Gaps 17;
Matches 80; Conservative 52; Mismatches 129;

Qy 116 KYNIMLRQIIANKEISKQO-----SIOEVTWVLLKAFFSIFREAHKSENLHPD 167
| | | | | : : : : : | | : | : : : : : | :
Db 2090 EVELIQYRKIKNSKERCESQKLEIDLLKSSKEELNLSIKATTOILELAKTKRMDELKTV 2149

Qy 168 NVIKKKNPSEKFKLIAEIC-----ICNEELN---VNPQDNG----- 202

Db 2150 NOLKKENERAOGKMKLLIKSCQLEEEKEIIOKELISOLAQAOEKOKGTVMYDKVDELTT 2209
QY 203 --ENISWTCORSSQOS-----IKSLAMRPRKRCGPGSLCCVQPRD----- 244
Db 2210 EIKELKTELEKTKADEYLDKYCSLLISHKLEKAKEMLETOVAHLCSQOSKODSRGSP 2269
QY 245 -LVPQVNVSAVASEGASPKP---WOLPSGVEPVGAKKSRIE-VWE-----PIPIRF 290
Db 2270 LIGPVVP-----GSPPIPSVTEKRLSSGONKASGKRQSSGIMWNGGPTPATPESF 2321
QY 291 QKTYGNPMWPRQKFAV--GVGSSWRTSA-----RVQKGNVGWEPHPRVPSGAPSS 339
Db 2322 SK-----KSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKKGFPAD-----IPTGKTSP 2367
QY 340 RAVRSPSSRLQKGRSTDLSL-----QHYPEKSTDP-----OCOPYKAGMESVYPK 386
Db 2368 YIIRRTTMATRTSPRLAOKLALSLPLSLGKENLAESSKPTAGGSRQKVAQSRSPVDSG 2427
QY 387 TVVAE-LTKTVGI 398
Db 2428 TILREPTTKSVPV 2440

RESULT 15
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORPMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 4.4%; Score 110; DB 1; Length 3248;
Best Local Similarity 21.4%; Pred. No. 0.45;

Matches 80; Conservative 52; Mismatches 129; Indels 112; Gaps 17;
QY 116 KVINMLMRQIILANKNEEISKOQ-----SIOEYTWVLKAFSFRREAHKSSENLHPD 167
Db 2818 EVELIQYTRKRLSKKECCSSOKLEIDLKSKKEELNNSLKATQIIEIKTKKMNLYV 2877
QY 168 NVIKKKNPSEGGFKFLAELIC-----ICNELN-----VNPQDNG----- 202
Db 2878 NOLKKENERAOGKMKLLIKSCQLEEEKEIIOKELISOLAQAOEKOKGTVMYDKVDELTT 2937
QY 203 --ENISWTCORSSQOS-----IKSLAMRPRKRCGPGSLCCVQPRD----- 244
Db 2938 EIKELKTELEKTKADEYLDKYCSLLISHKLEKAKEMLETOVAHLCSQOSKODSRGSP 2997
QY 245 -LVPQVNVSAVASEGASPKP---WOLPSGVEPVGAKKSRIE-VWE-----PIPIRF 290
Db 2998 LIGPVVP-----GSPPIPSVTEKRLSSGONKASGKRQSSGIMWNGGPTPATPESF 3049
QY 291 QKTYGNPMWPRQKFAV--GVGSSWRTSA-----RVQKGNVGWEPHPRVPSGAPSS 339
Db 3050 SK-----KSKKAVMSGIHPAEDTEGTEFEPEGLPEVVKKGFPAD-----IPTGKTSP 3095
QY 340 RAVRSPSSRLQKGRSTDLSL-----QHYPEKSTDP-----OCOPYKAGMESVYPK 386
Db 3096 YIIRRTTMATRTSPRLAOKLALSLPLSLGKENLAESSKPTAGGSRQKVAQSRSPVDSG 3155
QY 387 TVVAE-LTKTVGI 398
Db 3156 TILREPTTKSVPV 3168

Search completed: June 30, 2002, 11:53:23
Job time: 144 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2002, 11:49:04 ; Search time 31.16 Seconds

(without alignments)
1636.164 Million cell updates/sec

Title: US-09-997-610-2

Perfect score: 2494
Sequence: 1 IVYIPPLTAVIEHVEVACP.....GQFPPTGTAFTOCLYLHCW 459

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /A_Geneseq_032802.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	766	30.7	332	22	ABG02620
2	698	28.0	1984	22	ABG01338
3	694.5	27.8	273	22	ABG00491
4	676.5	27.1	939	22	ABG21982
5	665.5	26.7	901	22	ABG16683
6	653	26.2	220	22	ABG2981
7	628.5	25.2	333	22	ABG02511
8	602	24.1	466	22	ABG16460
9	573.5	23.0	270	22	ABG00481
10	570	22.9	439	22	ABG16421
11	570	22.9	439	22	ABG17512

12	540.5	21.7	261	22	ABG02502	Novel human diagno
13	540.5	21.7	261	22	ABG02632	Novel human diagno
14	528	21.2	766	22	ABG17044	Novel human diagno
15	523	21.0	636	22	ABG27632	Novel human diagno
16	487	19.5	561	22	ABG17236	Novel human diagno
17	486	19.5	290	22	ABG03735	Novel human diagno
18	483	19.4	595	22	ABG11326	Novel human diagno
19	477.5	19.1	239	22	ABG08795	Novel human diagno
20	476	19.1	533	22	ABG00280	Novel human diagno
21	475.5	19.1	310	22	ABG16947	Novel human diagno
22	473	19.0	224	22	ABG02622	Novel human diagno
23	462	18.5	289	22	ABG07801	Novel human diagno
24	459.5	18.4	553	22	ABG04234	Novel human diagno
25	458.5	18.3	297	22	ABG26263	Novel human diagno
26	457.5	18.3	842	22	ABG05486	Novel human diagno
27	435.5	17.5	528	22	ABG29732	Novel human diagno
28	434.5	17.4	276	22	ABG17722	Novel human diagno
29	424	17.0	204	22	ABG00278	Novel human diagno
30	417.5	16.7	1218	22	ABG06054	Novel human diagno
31	417.5	16.7	1218	22	ABG09637	Novel human diagno
32	417.5	16.7	1218	22	ABG10096	Novel human diagno
33	417.5	16.7	1218	22	ABG10796	Novel human diagno
34	417.5	16.7	1218	22	ABG14538	Novel human diagno
35	417.5	16.7	1218	22	ABG14753	Novel human diagno
36	417.5	16.7	1218	22	ABG17474	Novel human diagno
37	417.5	16.7	1218	22	ABG19905	Novel human diagno
38	417.5	16.7	1223	22	ABG08695	Novel human diagno
39	408.5	16.4	371	22	ABG05384	Novel human diagno
40	395	15.8	343	22	ABG16673	Novel human diagno
41	393	15.8	258	22	ABG00713	Novel human diagno
42	393	15.8	665	22	ABG03829	Novel human diagno
43	387	15.5	416	22	ABG17252	Novel human diagno
44	385	15.4	486	22	ABG11294	Novel human diagno
45	374.5	15.0	255	22	ABG17728	Novel human diagno

ALIGNMENTS

RESULT 1
ABG02620
ID ABG02620 standard; Protein: 332 AA.
XX
AC ABG02620;
XX
DF 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #2611.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-r2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
N-PSDB; AAS68807.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

Db 1766 pphvptcalspsgvtvrrgssrrpqrstkslhtpbgkadtqshpmkaartgspaaqj 1825
 Qy 384 -----PKTVVAELTKTY 396
 Db 1826 lcfargplgtlftawtaav 1844

RESULT 3
 ABG00491

ID ABG00491 standard; Protein; 273 AA.

AC ABG00491;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #482.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

XX N-PSDB; AAS64678.

PT New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biological activity

Claim 20; SEQ ID No 30850; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

Matches 141; Conservative 27; Mismatches 56; Indels 17; Gaps 3;

Qy 110 VELHCKVNIW--LMRKOILANKKEISKOOSIOEYTWLLKAFSPIREAKHSSENLHP 166

Db 28 lclgnkgrdlmwslerdldlyleefkqsglqemlwllkafsfvrgaekhslenlqp 87

Qy 167 DNVYKRRNPFSECKFKLAIEICINELVNPQDNGENISWTCQSSQSIKSLANPRR 226

Db 88 dsavekkipfseekfkpaaelcvseeppnhkpgrenenfracqkssqgppvpsaqgpr 147

Qy 227 -KWFCTGPGSLCCVQPRDLVPCVPVNS-----AVASRGASPKNQPLPSGYE 272

Db 148 kkwfcg1g1eslccv1srld1pcl1aasapamvkrgqdtfgamasegaahkpwllpgcve 207

Qy 273 PVGAKKSRLEWEPPIRFQKIGNPMWPMROKFAVGVGSSMRTSARVYQKGNVGEPPHRY 332

Db 208 pmngldqsgleweplprfgrmyknamwmsrgkfaagsepswtsaravlkgnvseppqts 267

Qy 333 P 333

Db 268 p 268

RESULT 4

ABG21982

ID ABG21982 standard; Protein; 939 AA.

AC ABG21982;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #21973.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

XX N-PSDB; AAS86169.

PT New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biological activity

Claim 20; SEQ ID No 52341; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

Query Match	26.2%	Score 653	DB 22	Length 220
Best Local Similarity	65.5%	Pred. No. 2.3e-49		
Matches 133	Conservative 15	Mismatches 49	Indels 6	Gaps 3
254	AVASGASGPKPMQQLSGVDPVCAKKSRIETVMEPPRIFFQKTYIGNPNKMPKPKFAVGVSSWR	313		
11	avaasgaspkpmqqlshgyevpvtlqtsrlvewpprfqtkmygnawmsrzkfaavagpswr	70		

OY	314	TSARVYQKNNWGEPEHRHPSGSPSSRAVRSPSSSLQGRSDSLQHPKEXTDPOCQ	373
Db	71	TSARVYQKNGVYGPETPRVYNGALSSGAVRPPSPITPQNGSTGDSJHAPYAAADTCQ	130
OY	374	PKVAGMESVPKYTVVAELTKVTGILILHCHDDVDRHGVRRDRFGALRFDCTPGEFRTYMG	433
Db	131	SVKAAEAREVPCOKATGAEALPKEMGPHLLHQrdInvrvvgdhlfgaIkKfdcpagfwcvmg	190
OY	434	P-VPLCFGQFFPEPTGTVAVFTQCLY 455	
Db	191	paapl----fwpm-splwngc1y 208	
RESULT 7			
ID	ABG02511	standard; Protein; 333 AA.	
XX	ABG02511;		
AC	ABG02511;		
XX			
DT	13-FEB-2002	(first entry)	
DE	Novel human diagnostic protein #2502.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX			
OS	Homo sapiens.		
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PE	30-MAR-2001; 2001WO-US06631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drimanac RT, Liu C, Tang YT;		
XX			
DR	WPI: 2001-639362/73.		
XX	N-PSDB: AAS66698.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
XX			
PS	Claim 20; SEQ ID NO 32870; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human		
CC	diagnostic amino acid sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp://wipo.int/pub/published_pct_sequences.		

XX	Sequence	333 AA:
Qy	Query Match	25.2%; Score 628.5; DB 22; Length 333;
Db	Best Local Similarity	60.1%; Pred. No. 5.9e-47;
	Matches 134; Conservative	15; Mismatches 55; Indels 19; Gaps
Qy	115	CKNIVW---LMKQILANKEETISKQOSTIOETWVWLKAFSPFREAEHKSSENLHPDWNVIK 171
Db	62	cpvdlwnfeleiddigylyveetkqgsitgnvtwllkafhfrfeehnlisenlhpadaae 121
Qy	172	KKNPESGKFKIAETICINELNVNPODNGENISWTQCRSSQSIKSLAMPRR-KWFC 230
Db	122	khspsfaekfkfpaaetclstneepv-----ngnvrsttcgrssqgfpqalrhrkkwfh 177
Qy	231	GIGGPGSLCCVQPPDLVPCVPVNS-----AVASEGASPKFWQLPSCGEVPGAKKS 279
Db	178	gvylygspccaaprlpvcvpatlpamaetgqcramasegavpklwqlphgilepasqks 237
Qy	280	RIEWEPPRIFFOKITGNGWMPROKFAVGVSSSMRTSARVYQG 322
Db	238	rlewepprifgmygnacvprdkfaaeaygsvrtaasapqy 280

CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. Abg00010-Abg3037 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 466 AA:
Query Match	24.1%; Score 602; DB 22; Length 466;
Best Local Similarity	63.4%; Pred. NO. 2.1e-44;
Matches 121; Conservative 14; Mismatches 32; Indels 24; Gaps	4;
QY	284 WEPP-----IREFOKITGNFWMRQKFAAGVSSWSRTSARVYQKGNG 325
Db	260 wpprdntcmckshngplrlhlkrlrfdkmgnaamprrqgfavgvsswrttsarmqgnyg 319
QY	336 WEPHRRVPSCAPSSRAVRKSPSSRLQGRSTDSLQHWPKEKSTDTQCQPYKKAAGMESVPE 385
Db	320 weiphryvtgapsgavrrgpssrpnqrstslhcypvgkaadtqpmkaagreaavpc 379
QY	386 KTVVAELTKTVGYTLHCHDLDVRHGVRKRDHGALRFDCPFGFTVMGP-VPLCFGQFPF 444
Db	380 kactlpekmetylllhgcoldvrhvxgdhlgallrtidcpagfqltcmapvpvl-----fwp 435
QY	445 FGTAVFPTQCLY 455
Db	436 i-spiwmgclly 445
RESULT 9	
ABG00481	
ID	ABG00481 standard; Protein: 270 AA.
XX	
AC	ABG00481;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #472.
XX	
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
PD	11-OCT-2001.
XX	
PJ	30-MAR-2001; 2001WO-US08631.
PF	
PR	31-MAR-2000; 2000US-0540217.
XX	
PA	23-AUG-2000; 2000US-0649167.
XX	
(HYSE-)	HYSEQ INC.
XX	
DR	Drmnac RT, Liu C, Tang YT;
XX	
WIPI:	2001-639362/73.
N-PSDB:	AAS64668.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
Claim 20:	SEQ ID No 30840; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 270 AA:

Query Match 23.0%; Score 573.5; DB 22; Length 270;
Best Local Similarity 48.3%; Pred. No. 3.1e-42;
Matches 128; Conservative 25; Mismatches 77; Indels 35; Gaps 5;

OY 202 GENIWTGORSQOSIKSLARPRKRWFCGTGPGSLCV-----QPRDLVPCVPVNS 253
DB 20 gddfflyleelskqsgsfxawe-----lgaicsshshgkrqggraw----- 61
OY 254 AVASEGAPKPMQLPSGVEPVGAKKSRIEWEPPRIFOKITGNPMPROKFAVGSSMR 313
DB 62 ataesaaskpqqfphsrepyagqskriewepptfqtmygnamspkfaagpblqr 121
OY 314 TSARVVGKNGWEPHARVPSGAPSSRAVRSPPSSRLQKGRSTDSLOHVPKSTDTQOQ 373
DB 122 tstrtlrkgvnelnppnrvtgklpbgavrrprepsfrangstnsllcvgtkdtctq 181
OY 374 PVKAGMSVPRKYVAELRTVGIYLLHCHDLDVRHGVRKRDHFGALKRDCPTGRTTWG 433
DB 182 pvtkaarraeapckatgaelpkrmtgthlhqcdldvrgvkgdhfgalftdcpagftwcg 241
OY 434 PVPLCFGQFPFPGTAVFT--OCLYL 456
DB 242 scs-----pflvlayfsylwvyl 259

RESULT 10
ABG16421
ID ABG16421 standard; Protein: 439 AA.

AC ABG16421;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16412.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.

XX
PN W0200175067-A2.
XX
XX 11-OCT-2001.

XX
PD 30-MAR-2001; 2001WO-US08631.
XX
PF 31-MAR-2000; 2000US-0540217.

XX
PR Human; chromosome mapping; gene mapping; gene therapy; forensic;

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
N-PSDB; AAS80608.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS Claim 20; SEQ ID No 46780; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 439 AA:

Query Match 22.9%; Score 570; DB 22; Length 439;
Best Local Similarity 62.2%; Pred. No. 1.2e-41;
Matches 117; Conservative 10; Mismatches 61; Indels 0; Gaps 0;

OY 257 SEGASPKPMQLPSGVEPVGAKKSRIEWEPPRIFOKITGNPMPROKFAVGSSMRTSA 316
DB 233 slsvsgrvpsasaagstalgksriewepvrrtqklygnsvtpwqkfaagvpswgtca 292
OY 317 RYVVGKNGWEPHARVPSGAPSSRAVRSPPSSRLQKGRSTDSLOHVPKSTDTQOQPYK 376
DB 293 rtvwkgivseppyrvtvgalpsgearrpppsrpnqsgndslcyvpgkstdtcqpyk 352
OY 377 AAGMSVPRKYVAELRTVGIYLLHCHDLDVRHGVRKRDHFGALKRDCPTGRTTWGPPY 436
DB 353 aagreglpcokatgvelsktighlhqcdldvrgvkgdhfgalftdcpagfritcmgpyt 412
OY 437 LCFGQFP 444
DB 413 plfwpisp 420

RESULT 11
ABG17512
ID ABG17512 standard; Protein: 439 AA.

AC ABG17512;
XX

DT 18-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #17503.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

Db 70 knpfseefkpaetctsnrepnvksprgwkmsprwt-rlhngslsh--hrprglgen 126
Qy 228 WFCGTGPGSLCCVQPRDLVPCVPVNSAVASEGASPKPMQLPSGVEPVGAKKSRIEWEPP 287
Db 127 gfvqgglgal-----hpnhsnbgfrgckpkpqlprdesasarkslievwmkpp 175
Qy 288 IRFOKITGNPMWMPROKFAVGVGSSWRTSARVYOKGNVGNEMPHRVPSCAPSSRAVRSP 347
Db 176 plfgkmygnvmpqlgtfaagagsswrtasarvqkgnvgwephrvtrtpspgalrrgpl 235
Qy 348 SSRLOKGRSTDSLQHVPEKST 368
Db 236 sarpqngsstslhvhvpgnat 256

RESULT 13
ABG02632
ID ABG02632 standard; Protein: 261 AA.
AC ABG02632;
XX 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #2623.
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PE 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX N-PSDB: AAS66819.
DR WPI: 2001-639362/73.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 32991; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 261 AA;
Query Match 21.7%; Score 540.5; DB 22; Length 261;
Best Local Similarity 48.7%; Pred. No. 2.4e-39;
Matches 127; Conservative 15; Mismatches 68; Indels 51; Gaps 7;
Qy 115 CKVNIW---LMRKQILANKKEISKOOSIOEVTWVLKAFSFIREAHHKSSENLAPDNVYK 171
Db 40 cprrdlwnfelerddlgylveelsksgsqge----- 69
Qy 172 KKNPSEGEFKIADICICNEELNV-NPDNGENISWTCORSSQOISISLANRPR---RK 227
Db 70 knpfseefkpaetctsnrepnvksprgwkmsprwt-rlhngslsh--hrprglgen 126
Qy 228 WFCGTGPGSLCCVQPRDLVPCVPVNSAVASEGASPKPMQLPSGVEPVGAKKSRIEWEPP 287
Db 127 gfvqgglgal-----hpnhsnbgfrgckpkpqlprdesasarkslievwmkpp 175
Qy 288 IRFOKITGNPMWMPROKFAVGVGSSWRTSARVYOKGNVGNEMPHRVPSCAPSSRAVRSP 347
Db 176 plfgkmygnvmpqlgtfaagagsswrtasarvqkgnvgwephrvtrtpspgalrrgpl 235
Qy 348 SSRLOKGRSTDSLQHVPEKST 368
Db 236 sarpqngsstslhvhvpgnat 256

RESULT 14
ABG17044
ID ABG17044 standard; Protein: 766 AA.
XX AC ABG17044;
XX 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #17035.
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PE 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX N-PSDB: AAS81231.
DR WPI: 2001-639362/73.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 47403; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

